

# Analysis of stochastic approximation schemes with set-valued maps in the absence of a stability guarantee and their stabilization

Vinayaka G. Yaji and Shalabh Bhatnagar,  
Department of Computer Science and Automation,  
Indian Institute of Science, Bangalore.  
vgyaji@gmail.com, shalabh@csa.iisc.ernet.in

January 27, 2017

## Abstract

In this paper, we analyze the behavior of stochastic approximation schemes with set-valued maps in the absence of a stability guarantee. We prove that after a large number of iterations if the stochastic approximation process enters the domain of attraction of an attracting set it gets locked into the attracting set with high probability. We demonstrate that the above result is an effective instrument for analyzing stochastic approximation schemes in the absence of a stability guarantee, by using it to obtain an alternate criteria for convergence in the presence of a locally attracting set for the mean field and by using it to show that a feedback mechanism, which involves resetting the iterates at regular time intervals, stabilizes the scheme when the mean field possesses a globally attracting set, thereby guaranteeing convergence. The results in this paper build on the works of *V.S. Borkar*, *C. Andrieu* and *H. F. Chen*, by allowing for the presence of set-valued drift functions.

## 1 Introduction

It is well known that several optimization and control tasks can be cast as a root finding problem. That is, given  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , one needs to find  $x^* \in \mathbb{R}^d$ , such that  $f(x^*) = 0$  (given such a point exists). Due to practical considerations, one usually has access to noisy measurements/estimations of the function whose root needs to be determined. An approach to solving such a problem with noisy measurements of  $f$ , is given by the recursion,

$$X_{n+1} - X_n - a(n)M_{n+1} = a(n)f(X_n), \quad (1)$$

where  $\{M_n\}_{n \geq 1}$ , denotes the noise arising in the measurement of  $f$  and having fixed an initial condition ( $X_0 \in \mathbb{R}^d$ ), the iterates  $\{X_n\}_{n \geq 1}$  are generated according to recursion (1). [1] under certain assumptions which include the Lipschitz continuity of the function  $f$ , boundedness of the iterates along almost every sample path (that is  $\mathbb{P}(\sup_{n \geq 0} \|X_n\| < \infty) = 1$ ) and a condition which ensures that the eventual contribution of the additive noise terms is negligible, showed that the linearly interpolated trajectory of recursion (1) tracks the flow of the ordinary differential equation (o.d.e.) given by,

$$\frac{dx}{dt} = f(x). \quad (2)$$

Such a trajectory is called an *asymptotic pseudotrajectory* for the flow of o.d.e.(2) (for a precise definition see [1]). Suppose the set of zeros of  $f$  is a globally asymptotically stable set for the flow of o.d.e. (2), then it was shown that the limit set of an asymptotic pseudotrajectory was contained in such a set and hence the iterates  $\{X_n\}_{n \geq 0}$  converge in the limit to a root of the function  $f$ .

In order to analyze recursion (1) when the function  $f$  is no longer Lipschitz continuous or even continuous, but is just measurable satisfying the linear growth property, that is for every  $x \in \mathbb{R}^d$ ,  $\|f(x)\| \leq K(1 + \|x\|)$  for some  $K > 0$ , or when there is a non-additive noise/control component taking values in a compact set whose law is not known (in which case the recursion (1) takes the form  $X_{n+1} - X_n - a(n)M_{n+1} = a(n)f(X_n, U_n)$ , where  $U_n$  denotes the noise/control), the above mentioned o.d.e.

method needed to be extended to recursions with much weaker requirements on the function  $f$ . This was accomplished in [2], where the asymptotic behavior of the recursion given by,

$$X_{n+1} - X_n - a(n)M_{n+1} \in a(n)F(X_n), \quad (3)$$

was studied, where  $F$  is a set-valued map satisfying some conditions (while the other quantities have same interpretation as in (1)). Under the assumption of stability of iterates (that is  $\mathbb{P}(\sup_{n \geq 0} \|X_n\| < \infty) = 1$ ) and appropriate conditions on the additive noise terms, in [2], it was shown that the linearly interpolated trajectory of recursion (3) tracks the flow of the differential inclusion (d.i.) given by,

$$\frac{dx}{dt} \in F(x). \quad (4)$$

We refer the reader to [3, Ch. 5.3] for a detailed argument as to how the measurable case and the case with unknown noise/control be recast in the form of recursion (3). For a brief summary of the convergence analysis of recursion (3) we refer the reader to section 3.1 of this paper.

Common to the analysis of both recursion (1) and (3) is the assumption on the stability of the iterates, that is  $\mathbb{P}(\sup_{n \geq 0} \|X_n\| < \infty) = 1$ . The condition of stability is highly non-trivial and difficult to verify. Over the years significant effort has gone into providing sufficient conditions for stability (see [4], [5]). In [6], it was shown that for recursion (1), in the absence of stability guarantee, the probability of converging to an attracting set of o.d.e. (2) given that the iterates lie in a neighborhood of it converged to one as the index ( $n$ ) in which the iterate entered the neighborhood of the attracting set increased to infinity. This probability of the iterates converging to an attracting set given that the iterate lies in a neighborhood of it is called the *lock-in probability* and in [6] a lower bound for the same was used to obtain sample complexity bounds for recursion (1). Further a tighter lower bound for the lock-in probability was derived in [7] under a slightly stronger noise assumption and used to obtain convergence guarantee when the law of the iterates are tight. In this paper we extend the results in [6] to the case of stochastic approximation schemes with set-valued maps as in recursion (3).

## 1.1 Contributions and organization of the paper

We first provide a lower bound for the lock-in probability of stochastic approximation schemes with set-valued maps as in recursion (3). The bound is derived under an assumption on the additive noise terms which is stronger than the corresponding in [6], which is necessitated due to the lack of Lipschitz continuity of the drift function  $F$ . We establish that,

$$\mathbb{P}(X_n \rightarrow A \text{ as } n \rightarrow \infty | X_{n_0} \in \mathcal{O}') \geq 1 - 2de^{-\tilde{K}/b(n_0)},$$

for  $n_0$  large, where,  $A \subseteq \mathbb{R}^d$ , denotes an attracting set of DI 4,  $\mathcal{O}'$  is an open neighborhood of  $A$  with compact closure,  $\tilde{K}$  is some positive constant and  $\{b(n)\}_{n \geq 0}$  is a sequence of reals converging to zero, which are step size dependent.

Having summarized the convergence analysis under stability in section 3.1, we state the lock-in probability bound in section 3.2 and provide a few implications of the same. Using the lock-in probability result we provide an alternate criteria for convergence in the presence of a locally attracting set which removes the need to verify stability. A detailed comparison between the obtained convergence guarantee and the corresponding in the presence of stability is also provided.

Proof of the lock-in probability result is presented in section 5. The proof relies heavily on the insights obtained from the analysis in [6] for single-valued maps. From the analysis in [6], it is evident that the Lipschitz continuity of the drift function  $f$  plays a crucial role in obtaining events and decoupling error contributions which in turn are necessary to obtain the bound in the inequality above. But in the recursion studied in this paper (that is recursion (3)), the drift function  $F$  is set-valued and the assumptions under which we study the said recursion (which are summarized in section 2), the drift function  $F$  is not even continuous. We overcome this problem by first obtaining a sequence of locally Lipschitz continuous set-valued maps which approximate the drift function  $F$  from above and then parameterizing them using the Stienen selection procedure. The associated results are summarized in section 5.1. This enables us to write recursion (3) in the form of recursion (1), but with locally Lipschitz continuous drift functions. Further the relation between the solutions of differential inclusions with the approximating set-valued maps as their vector field and those of DI (4), is established in section 5.2. Having written recursion (3) in the form of recursion (1), we then collect sample paths of interest in section 5.3. Along the sample paths that are collected the iterates are such that, having entered a neighborhood

of the attracting set at iteration  $n_0$ , the iterates will infinitely often enter the said neighborhood and the time elapsed between successive visits to the neighborhood of the attracting set can be upper bounded by a constant which is mean field dependent. Further we show that the probability of occurrence of such sample paths can be lower bounded by error contributions due to additive noise terms alone after a large number of iterations. Using the concentration inequality for martingale sequences we obtain the lock-in probability bound in section 5.4.

Using the lock-in probability result we design a feedback mechanism which enables us to stabilize the stochastic approximation scheme in the presence of a globally attracting set for DI (4). The feedback mechanism involves resetting the iterates at regular time intervals if they are found to be lying outside a certain compact set. This approach to stabilization has been studied in various forms for stochastic approximation schemes with single-valued drift functions as in recursion (1), in [8], [9], [10] and [11] to name a few. We extend the same to the case of set-valued drift functions. The main idea in the analysis of such a scheme is to show that along almost every sample path of the modified recursion, the number of resets that are performed is finite, thereby guaranteeing that eventually the iterates lie within a compact set. We observe that the lock-in probability result (to be precise the approach adopted to obtain the lock-in probability result) plays a central role in showing that the number of resets performed remain finite. Having shown that the iterates eventually lie within a compact set, we use the convergence arguments from [2] to argue that the iterates generated by the modified scheme converge to the globally attracting set of DI (4). The modified scheme is presented and explained in detail in section 4. The proof of the finite resets theorem is presented in section 6. The procedure employed to collect sample paths in the proof of the lock-in probability result can be used to collect sample paths where only finite number of resets have occurred in the modified scheme and this in turn enables us show that the number of resets are finite almost surely.

Finally, we conclude by providing a few directions for future work in section 7.

## 2 Recursion and assumptions

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and  $\{X_n\}_{n \geq 0}$  be a sequence of  $\mathbb{R}^d$ -valued random variables on  $\Omega$ , such that for every  $n \geq 0$ ,

$$X_{n+1} - X_n - a(n)M_{n+1} \in a(n)F(X_n), \quad (5)$$

where,

(A1)  $F : \mathbb{R}^d \rightarrow \{\text{subsets of } \mathbb{R}^d\}$  is a set-valued map which for every  $x \in \mathbb{R}^d$  satisfies the following:

- (i)  $F(x)$  is a convex and compact subset of  $\mathbb{R}^d$ ,
- (ii) there exists  $K > 0$  (independent of  $x$ ) such that  $\sup_{y \in F(x)} \|y\| \leq K(1 + \|x\|)$ ,
- (iii) for every  $\mathbb{R}^d$ -valued sequence  $\{x_n\}_{n \geq 1}$  converging to  $x$  and for every sequence  $\{y_n \in F(x_n)\}_{n \geq 1}$  converging to  $y \in \mathbb{R}^d$ , we have that  $y \in F(x)$ .

(A2)  $\{a(n)\}_{n \geq 0}$  is a sequence of positive real numbers satisfying,

- (i)  $\sum_{n=0}^{\infty} a(n) = \infty$ ,
- (ii)  $\sum_{n=0}^{\infty} (a(n))^2 < \infty$ .

(A3)  $\{M_n\}_{n \geq 1}$  is a  $\mathbb{R}^d$ -valued, martingale difference sequence with respect to the filtration  $\{\mathcal{F}_n := \sigma(X_m, M_m, m \leq n)\}$ . Furthermore,  $\{M_n\}_{n \geq 1}$  are such that,

$$\|M_{n+1}\| \leq K(1 + \|x_n\|) \text{ a.s.,}$$

for every  $n \geq 0$ , for some constant  $K > 0$ .

Assumption (A1) ensures that the set-valued map  $F$  is a Marchaud map. The condition (A1)(ii) is called the *linear growth property* since it ensures that the size of the sets  $F(x)$  grow linearly with respect to the distance from the origin. The condition (A1)(iii) is called the *closed graph property* since it states that the graph of the set-valued map  $F$ , defined as,

$$\{(x, y) \in \mathbb{R}^{2d} : x \in \mathbb{R}^d, y \in F(x)\},$$

is a closed subset of  $\mathbb{R}^{2d}$ . The map  $F$  being a Marchaud map ensures that the differential inclusion (DI) given by,

$$\frac{dx}{dt} \in F(x), \quad (6)$$

possesses at least one solution through every initial condition. By a solution of DI (6) with initial condition  $x_0 \in \mathbb{R}^d$ , we mean an absolutely continuous function  $\mathbf{x} : \mathbb{R} \rightarrow \mathbb{R}^d$  such that  $\mathbf{x}(0) = x_0$  and for almost every  $t \in \mathbb{R}$ ,  $\frac{d\mathbf{x}(t)}{dt} \in F(\mathbf{x}(t))$ . DI (6) is the *mean field* of recursion (5) and its dynamics play an important role in describing the asymptotic behavior of recursion (5).

Assumption (A2) states the conditions to be satisfied by the step size sequence  $\{a(n)\}_{n \geq 0}$ . Square summability (that is (A2)(ii)) is needed later in the analysis for obtaining a probability bound on certain tail events associated with the additive noise terms  $\{M_n\}_{n \geq 1}$ .

Assumption (A3), defines the *martingale noise model*. These terms denote the noise arising in the measurement of  $F(\cdot)$ . This condition holds in several reinforcement learning applications (see [3, Ch. 10])

**Remark** Clearly when  $\{M_n\}_{n \geq 1}$  are i.i.d. zero mean and bounded, assumption (A3) is satisfied. Further, since the drift function in recursion (5) is a set-valued map, scenarios where the measurement noise terms possess a bounded bias can be recast in the form of recursion (5) as explained below.

Consider the recursion given by,

$$X_{n+1} - X_n - a(n)M_{n+1} - a(n)\eta_{n+1} = a(n)f(X_n), \quad n \geq 0, \quad (7)$$

where  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a single-valued Lipschitz continuous map, for every  $n \geq 0$ ,  $\eta_{n+1}$  denotes the bias in the measurement noise. Let the bias terms  $\{\eta_n\}_{n \geq 1}$  be bounded by a positive constant, say  $\epsilon > 0$  (that is, for every  $n \geq 1$ ,  $\|\eta_n\| \leq \epsilon$ ). Then, recursion (7) can be written in the form of recursion (5) with set-valued map  $F$ , given by,  $F(x) = \{f(x) + \eta : \|\eta\| \leq \epsilon\}$ , for every  $x \in \mathbb{R}^d$ . We refer the reader to [3, ch. 5.3] for several other variants of the standard stochastic approximation scheme which can be analyzed with the help of recursion (5).

### 3 Lock-in probability for stochastic recursive inclusions

In order to state the main result of this paper, definition of the flow of a DI, an attracting set for such a dynamical system are needed. We recall these notions below and we state them with respect to the mean field of recursion (5) (for a detailed description and associated results see [2]).

The *flow* of DI (6) is given by the set-valued map  $\Phi : \mathbb{R} \times \mathbb{R}^d \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ , where for every  $(t, x) \in \mathbb{R} \times \mathbb{R}^d$ ,

$$\Phi(t, x) := \{\mathbf{x}(t) \in \mathbb{R}^d : \mathbf{x}(\cdot) \text{ is a solution of DI (6) with } \mathbf{x}(0) = x\}. \quad (8)$$

A compact set  $A \subset \mathbb{R}^d$  is an *attracting set* for the flow of DI (6), if there exists an open neighborhood of  $A$ , say  $\mathcal{O}$ , with the property that for every  $\epsilon > 0$ , there exists a time  $T > 0$  (depending on  $\epsilon$  and  $\mathcal{O}$ ) such that for every  $t \geq T$  and for every  $x \in \mathcal{O}$ ,  $\Phi(t, x) \in N^\epsilon(A)$ , where  $N^\epsilon(A)$  denotes the  $\epsilon$ -neighborhood of  $A$ . Such a neighborhood  $\mathcal{O}$  of an attracting set  $A$  is called the *fundamental neighborhood* of  $A$ .

The set of initial conditions in  $\mathbb{R}^d$  from which the flow is attracted to an attracting set  $A$  is called the *basin of attraction* and is denoted by  $B(A)$ . Formally,

$$B(A) := \left\{x \in \mathbb{R}^d : \bigcap_{t \geq 0} \overline{\{\Phi(q, x) : q \geq t\}} \subseteq A\right\}.$$

An attracting set  $A$  is said to be *globally attracting* if,  $B(A) = \mathbb{R}^d$ .

#### 3.1 Summary of the asymptotic analysis under stability

Let  $t(0) := 0$  and for every  $n \geq 1$ ,  $t(n) := \sum_{k=0}^{n-1} a(k)$ . The linearly interpolated trajectory of recursion (5), is given by the stochastic process  $\bar{X} : \Omega \times \mathbb{R} \rightarrow \mathbb{R}^d$ , where for every  $(\omega, t) \in \Omega \times [0, \infty)$ ,

$$\bar{X}(\omega, t) := \left(\frac{t - t(n)}{t(n+1) - t(n)}\right) X_{n+1}(\omega) + \left(\frac{t(n+1) - t}{t(n+1) - t(n)}\right) X_n(\omega), \quad (9)$$

where  $n$  is such that  $t \in [t(n), t(n+1))$  and for every  $(\omega, t) \in \Omega \times (-\infty, 0)$ ,  $\bar{X}(\omega, t) := X_0(\omega)$ .

For  $\omega \in \Omega$ , the limit set map of  $\bar{X}$  is given by,  $\lambda : \Omega \rightarrow \{\text{subsets of } \mathbb{R}^d\}$  where for every  $\omega \in \Omega$ ,

$$\lambda(\omega) := \bigcap_{t \geq 0} \overline{\{\bar{X}(\omega, q) : q \geq t\}}. \quad (10)$$

In [2], under assumptions (A1) – (A3) along with the additional assumption of stability of the iterates (that is  $\mathbb{P}(\sup_{n \geq 0} \|X_n\| < \infty) = 1$ ), it was shown that for almost every  $\omega \in \Omega$ , the linearly interpolated trajectory of recursion (5),  $\bar{X}(\omega, \cdot)$ , is an *asymptotic pseudotrajectory* for the flow of DI (6). More precisely, for almost every  $\omega \in \Omega$ ,  $\bar{X}(\omega, \cdot)$  was shown to satisfy the following:

- (a) The family of shifted trajectories given by  $\{\bar{X}(\omega, \cdot + t)\}_{t \geq 0}$  is relatively compact in  $\mathcal{C}(\mathbb{R}, \mathbb{R}^d)$  where  $\mathcal{C}(\mathbb{R}, \mathbb{R}^d)$  denotes the metric space of all continuous functions on  $\mathbb{R}$  taking values in  $\mathbb{R}^d$  with metric  $\mathbf{D}$ , which for every  $\mathbf{z}, \mathbf{z}' \in \mathcal{C}(\mathbb{R}, \mathbb{R}^d)$  is given by

$$\mathbf{D}(\mathbf{z}, \mathbf{z}') = \sum_{k=1}^{\infty} \frac{1}{2^k} \min\{\|\mathbf{z} - \mathbf{z}'\|_{[-k, k]}, 1\}, \quad (11)$$

where  $\|\mathbf{z} - \mathbf{z}'\|_{[-k, k]} := \sup_{t \in [-k, k]} \|\mathbf{z}(t) - \mathbf{z}'(t)\|$ .

- (b) Every limit point of the shifted trajectories  $\{\bar{X}(\omega, \cdot + t)\}_{t \geq 0}$  is a solution of the DI (6).

From [2, Thm. 4.3], it follows that for almost every  $\omega \in \Omega$ , the limit set of the linearly interpolated trajectory  $\bar{X}(\omega, \cdot)$ ,  $\lambda(\omega)$ , is a non-empty, compact and an internally chain transitive (ICT) set for the flow of DI (6) (see [2, Defn. VI] for definition of an ICT set). Now using [2, Thm. 3.23] the main convergence result of [2] follows and is stated below.

**Theorem 3.1** *Let  $A \subseteq \mathbb{R}^d$  be an attracting set for the flow of DI (6). Under assumptions (A1) – (A3),*

- (a) *for almost every  $\omega \in \{\omega \in \Omega : \sup_{n \geq 0} \|X_n(\omega)\| < \infty\} \cap \{\omega \in \Omega : \lambda(\omega) \cap B(A) \neq \emptyset\}$ ,  $\lambda(\omega) \subseteq A$  and therefore as  $n \rightarrow \infty$ ,  $X_n(\omega) \rightarrow A$ .*
- (b) *if  $B(A) = \mathbb{R}^d$  (that is  $A$  is a globally attracting set), then for almost every  $\omega \in \{\omega \in \Omega : \sup_{n \geq 0} \|X_n(\omega)\| < \infty\}$ ,  $\lambda(\omega) \subseteq A$  and therefore as  $n \rightarrow \infty$ ,  $X_n(\omega) \rightarrow A$ .*

The assumption of stability of the iterates used to obtain the above convergence result is highly non-trivial and difficult to verify. Moreover the proof method used to prove the above convergence result cannot be modified in a straight forward manner to obtain a similar convergence guarantee. This warrants an alternate approach to study the behavior of recursion (5) in the absence of stability guarantee and we accomplish this by extending the lock-in probability result from [6] to the set-valued case. Using the obtained lock-in probability bound we recover convergence guarantee similar to Theorem 3.1 while eliminating the need to verify stability.

## 3.2 Main result and its implications

Before we state the main result, we state an assumption which fixes the attracting set of interest.

- (A4) Let  $A \subseteq \mathbb{R}^d$ , be an attracting set of DI (6) (the mean field of recursion (5)) with  $\mathcal{O} \subseteq \mathbb{R}^d$  as its fundamental neighborhood of attraction.

Let  $\mathcal{O}'$  be an open neighborhood of the attracting set  $A$  (as in (A4)) such that  $\bar{\mathcal{O}}'$  is compact and  $\bar{\mathcal{O}}' \subseteq \mathcal{O}$ . Then the main result of the paper can be stated as follows.

**Theorem 3.2** (Lock-in probability) *Under assumptions (A1) – (A4), there exists a constant  $\tilde{K} > 0$  (depending on the attracting set  $A$  and  $\mathcal{O}'$ ) and an  $N_0 \geq 1$  such that, for every  $n_0 \geq N_0$ , for every  $E \in \mathcal{F}_{n_0}$  satisfying  $E \subseteq \{\omega \in \Omega : X_{n_0}(\omega) \in \mathcal{O}'\}$  and  $\mathbb{P}(E) > 0$ , we have that,*

$$\mathbb{P}(X_n \rightarrow A \text{ as } n \rightarrow \infty | E) \geq 1 - 2de^{-\tilde{K}/b(n_0)},$$

where, for every  $n \geq 0$ ,  $b(n) := \sum_{k=n}^{\infty} (a(k))^2$ .

There are two immediate implications of the above result and are stated below, one of which serves as an alternate convergence result in the absence of stability guarantees, that is it allows us to obtain the convergence guarantee in Theorem 3.1(a) without the need to verify whether a given sample path satisfies  $\sup_{n \geq 0} \|X_n(\omega)\| < \infty$ .

- (1) As a consequence of assumption (A2)(ii), we have that  $\lim_{n \rightarrow \infty} b(n) = 0$ . Therefore from Theorem 3.2, if the observation that iterate lies in a neighborhood of the attracting set is made later in time ( $n_0$ ), the probability of converging to the attracting set increases and converges to one as  $n_0 \rightarrow \infty$ . Formally,

$$\lim_{n_0 \rightarrow \infty} \mathbb{P}(X_n \rightarrow A \text{ as } n \rightarrow \infty | X_{n_0} \in \mathcal{O}') = 1.$$

- (2) Suppose  $\mathbb{P}(\cap_{N \geq 0} \cup_{n \geq N} \{X_n \in \mathcal{O}'\}) > 0$  (if  $\mathbb{P}(\cap_{N \geq 0} \cup_{n \geq N} \{X_n \in \mathcal{O}'\}) = 0$  then the iterates almost surely do not converge to the attracting set  $A$ ). Then for every  $N \geq 0$ ,  $\mathbb{P}(\cup_{n \geq N} \{X_n \in \mathcal{O}'\}) > 0$  and

$$\cup_{n \geq N} \{X_n \in \mathcal{O}'\} = \{X_N \in \mathcal{O}'\} \cup (\cup_{n > N} \{X_k \notin \mathcal{O}', \text{ for } N \leq k \leq n-1, X_n \in \mathcal{O}'\}),$$

where, the union in the R.H.S. is disjoint. Then by Theorem 3.2, for every  $N \geq N_0$ ,

$$\begin{aligned} & \mathbb{P}(\{X_n \rightarrow A \text{ as } n \rightarrow \infty\} \cap (\cap_{N \geq 0} \cup_{n \geq N} \{X_n \in \mathcal{O}'\})) \\ & \geq \sum_{n \geq N} \mathbb{P}(\{X_n \rightarrow A \text{ as } n \rightarrow \infty\} \cap \{X_k \notin \mathcal{O}', \text{ for } N \leq k \leq n-1, X_n \in \mathcal{O}'\}) \\ & = \sum_{n \geq N} \left[ \mathbb{P}(\{X_n \rightarrow A \text{ as } n \rightarrow \infty | \{X_k \notin \mathcal{O}', \text{ for } N \leq k \leq n-1, X_n \in \mathcal{O}'\}) \right. \\ & \quad \left. \mathbb{P}(\{X_k \notin \mathcal{O}', \text{ for } N \leq k \leq n-1, X_n \in \mathcal{O}'\}) \right] \\ & \geq \sum_{n \geq N} \left(1 - 2de^{-\tilde{K}/b(n)}\right) \mathbb{P}(\{X_k \notin \mathcal{O}', \text{ for } N \leq k \leq n-1, X_n \in \mathcal{O}'\}) \\ & \geq \left(1 - 2de^{-\tilde{K}/b(N)}\right) \sum_{n \geq N} \mathbb{P}(\{X_k \notin \mathcal{O}', \text{ for } N \leq k \leq n-1, X_n \in \mathcal{O}'\}) \\ & = \left(1 - 2de^{-\tilde{K}/b(N)}\right) \mathbb{P}(\cup_{n \geq N} \{X_n \in \mathcal{O}'\}) \\ & \geq \left(1 - 2de^{-\tilde{K}/b(N)}\right) \mathbb{P}(\cap_{N \geq 0} \cup_{n \geq N} \{X_n \in \mathcal{O}'\}). \end{aligned}$$

The above inequality is true for every  $N \geq N_0$ . Taking limit and using the fact that  $\lim_{n \rightarrow \infty} b(n) = 0$ , we get that,

$$\mathbb{P}(\{X_n \rightarrow A \text{ as } n \rightarrow \infty\} \cap (\cap_{N \geq 0} \cup_{n \geq N} \{X_n \in \mathcal{O}'\})) = \mathbb{P}(\cap_{N \geq 0} \cup_{n \geq N} \{X_n \in \mathcal{O}'\}).$$

Therefore from the above we can conclude that,

**Corollary 3.3** *Under assumptions (A1) – (A4), for almost every  $\omega \in \cap_{N \geq 0} \cup_{n \geq N} \{X_n \in \mathcal{O}'\}$ ,  $X_n(\omega) \rightarrow A$  as  $n \rightarrow \infty$ .*

**Remark** In comparison with Theorem 3.1(a), the condition that  $\omega \in \cap_{N \geq 0} \cup_{n \geq N} \{X_n \in \mathcal{O}'\}$  is stronger than the requirement that  $\omega \in \{\lambda(\omega) \cap B(A) \neq \emptyset\}$  because the former requires the iterate sequence to enter an open neighborhood of  $A$  with compact closure infinitely often while the latter requires the iterates to enter the basin of attraction of  $A$  infinitely often which is larger than  $\mathcal{O}'$ . But in the presence of stability we have that,

$$\{\sup_{n \geq 0} \|X_n\| < \infty\} \cap \{\lambda(\cdot) \cap B(A) \neq \emptyset\} \subseteq \cap_{N \geq 0} \cup_{n \geq N} \{X_n \in \mathcal{O}'\}.$$

Further, as a consequence of Corollary 3.3, we have that,

$$\mathbb{P}(\{\sup_{n \geq 0} \|X_n\| < \infty\} \cap \{\lambda(\cdot) \cap B(A) \neq \emptyset\}) = \mathbb{P}(\cap_{N \geq 0} \cup_{n \geq N} \{X_n \in \mathcal{O}'\}),$$

or in other words, the sample paths which visit  $\mathcal{O}'$  infinitely often and are unstable, occur with zero probability.

## 4 Application: Stabilization via resetting

In this section we modify recursion (5) in such a way that the modified procedure yields sample paths which are stable (that is lie in a compact set almost surely) which in turn allows us to recover the convergence result as in Theorem 3.1(b) without the need to verify stability, in the presence of a globally attracting set for the mean field. That is, we replace assumption (A4) with the following stronger requirement.

(A4)' Let  $A \subseteq \mathbb{R}^d$  be a globally attracting set for the flow of DI (6).

The modification that we propose involves resetting the iterates at regular time intervals if they are found to be lying outside a certain compact set. Let the initial condition  $X_0(\omega) = x_0 \in \mathbb{R}^d$  for every  $\omega \in \Omega$  and  $\{r_n \in (0, \infty)\}_{n \geq 0}$  be such that,

- (1)  $\|x_0\| < r_0$ ,
- (2) for every  $n \geq 0$ ,  $r_n < r_{n+1}$ ,
- (3)  $\lim_{n \rightarrow \infty} r_n = \infty$ .

The modified scheme, henceforth referred to as *stabilized stochastic recursive inclusion* (SSRI) is where every sample path is generated as outlined in Algorithm 1.

---

**Algorithm 1** SSRI given  $x_0$  and  $\{r_k\}_{k \geq 0}$

---

$n \leftarrow 0$ $k \leftarrow 0$ $t_e \leftarrow 0$ $T_W > 0$ $n_W \leftarrow 1$ $X'_0(\omega) \leftarrow x_0$ <b>while</b> $n \geq 0$ <b>do</b> $X_{n+1}(\omega) - X'_n(\omega) - a(n)M_{n+1}(\omega) \in a(n)F(X'_n(\omega))$ $t_e \leftarrow t_e + a(n)$ <b>if</b> $t_e \geq T_W$ <b>then</b> <b>if</b> $n_W = 1$ <b>then</b> <b>if</b> $\ X_{n+1}(\omega)\  > r_k$ <b>then</b> $X'_{n+1}(\omega) = x_0$ $k \leftarrow k + 1$ <b>else</b> $X'_{n+1}(\omega) \leftarrow X_{n+1}(\omega)$ <b>end if</b> $n_W = 2^k$ <b>else</b> $n_W \leftarrow n_W - 1$ $X'_{n+1}(\omega) = X_{n+1}(\omega)$ <b>end if</b> $t_e \leftarrow 0$ <b>else</b> $X'_{n+1}(\omega) \leftarrow X_{n+1}(\omega)$ <b>end if</b> $n \leftarrow n + 1$ <b>end while</b>	▷ Initialize iteration count ▷ Initialize reset count ▷ Initialize time elapsed since last check ▷ Initialize window length ▷ Initialize window count ▷ Initialize initial condition ▷ Obtain $X_{n+1}(\omega)$ ▷ Update the time elapsed ▷ Is time elapsed greater than the window length? ▷ Have sufficient number of windows elapsed? ▷ Is the iterate lying outside a compact set? ▷ Perform reset ▷ Increment reset count ▷ Perform no reset ▷ Reset window count ▷ Decrement window count ▷ Perform no reset ▷ Reset time elapsed ▷ Perform no reset ▷ Increment iteration count
---	---

---

A flowchart depicting the flow of control in Algorithm 1 is presented in Figure 1. In order to understand the algorithm let us consider the scenario where the  $k^{th}$  reset has been performed at iteration index  $n_0$ . Then the algorithm checks whether the iterate lies in the compact set  $r_k U$  (closed ball of radius  $r_k$  centered at the origin) after approximately  $2^k T_W$  amount of time has elapsed (for the relation between time and iteration index see section 3.1). Now either a reset occurs or the iterate is left unchanged.

- (a) If the iterate is left unchanged then the next reset check is performed after  $2^k T_W$  amount of time has elapsed.

(b) If the iterate is reset, then, the next check is performed after  $2^{k+1}T_W$  amount of time has elapsed.

In fact it would suffice if the time between successive reset checks were set to be greater than a certain threshold which is determined by the minimum time needed by the flow of the mean field (that is DI (6)) to reach the attracting set  $A$  from any initial condition in a compact neighborhood of it. But in practical scenarios one may not be able to compute such a time and hence may not be able to determine the required threshold. This approach of increasing time duration between successive reset checks with increasing reset count allows us to bypass this problem. The choice of exponentially increasing durations is one of convenience as it simplifies notations involved in proving certain results later.

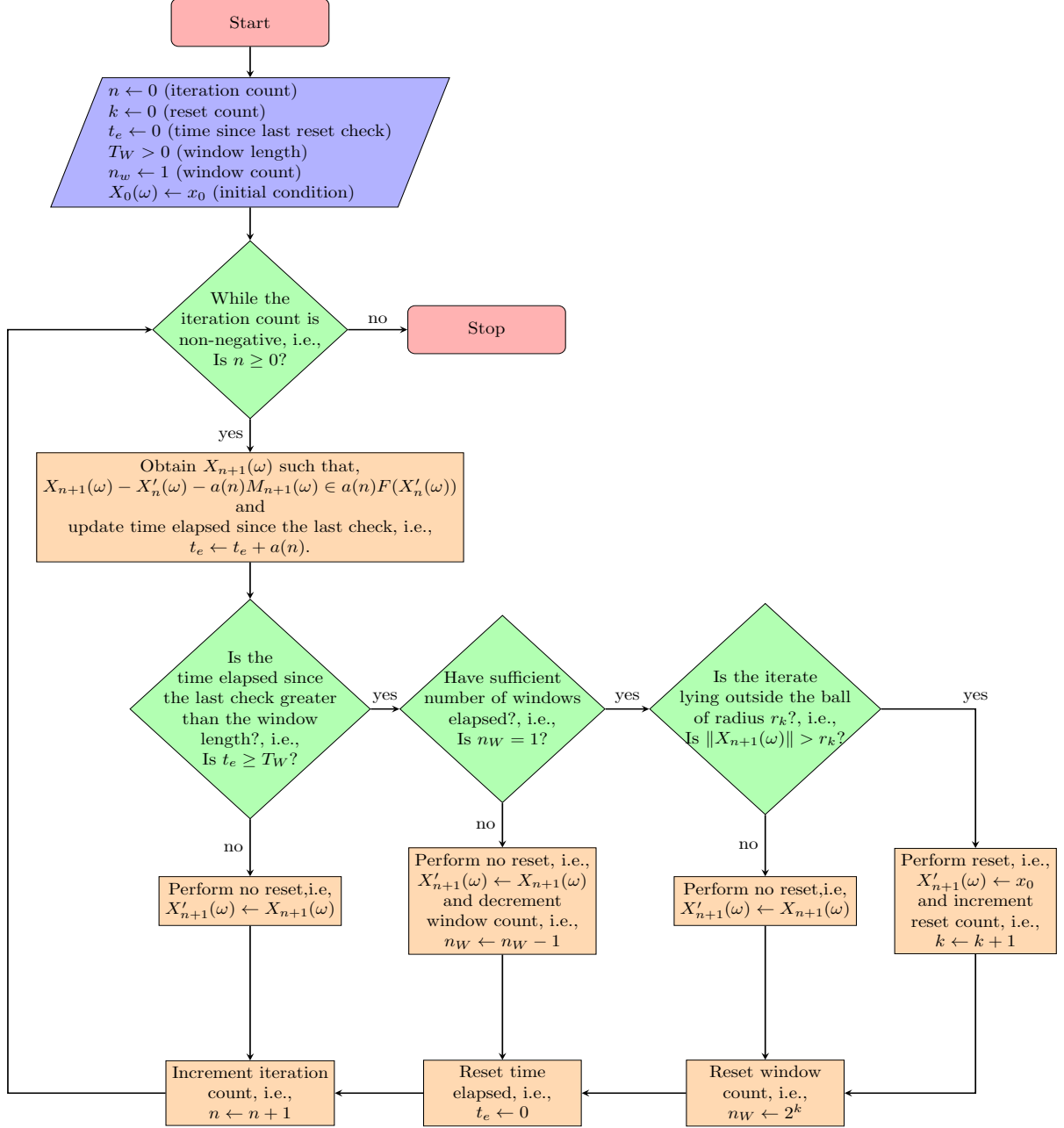


Figure 1: Flowchart depicting the flow of control in Algorithm 1

For every  $n \geq 1$ , define the indicator random variable  $\chi_n : \Omega \rightarrow \{0, 1\}$  such that, for every  $\omega \in \Omega$ ,

$$\chi_n(\omega) = \begin{cases} 0 & \text{if } X_n(\omega) = X'_n(\omega), \\ 1 & \text{if } X_n(\omega) \neq X'_n(\omega). \end{cases} \quad (12)$$

We assume that the noise terms  $\{M_n\}_{n \geq 1}$  satisfy the following version of assumption (A3).



(A3)'  $\{M_n\}$  is a martingale difference sequence with respect to the filtration  $\{\mathcal{F}_n\}_{n \geq 1}$ , where, for every  $n \geq 1$ ,  $\mathcal{F}_n$  denotes the smallest  $\sigma$ -algebra generated by the iterates  $X_m$  (that is the iterates before the reset operation) and noise terms  $M_m$ , for  $0 \leq m \leq n$  (then it is easy to show that for every  $n \geq 1$ ,  $X'_n$  and hence  $\chi_n$  are  $\mathcal{F}_n$  measurable). Since for every  $n \geq 1$ ,  $M_n$  denotes the noise arising in the estimation (or measurement) of  $F$  at  $X'_{n-1}$ , we assume that the energy of the noise depends on  $X'_{n-1}$ . That is for every  $n \geq 0$ ,  $\|M_{n+1}\| \leq K(1 + \|X'_n\|)$  a.s..

The next theorem says that, for almost every sample path generated by Algorithm 1, the total number of resets is finite, thereby guaranteeing stability. The proof of this theorem (provided in section 6) crucially hinges on a lower bound for the probability of the event that there are no future resets given that there are a certain number of resets up until iteration  $n_0$  for some large  $n_0$ . Specifically it requires the probability of the above mentioned event to converge to one as  $n_0$  tends to infinity and this is guaranteed by Theorem 3.2.

**Theorem 4.1** (Finite resets) *Under assumptions (A1), (A2), (A3)' and (A4)',  $\mathbb{P}(\{\omega \in \Omega : \sum_{n=1}^{\infty} \chi_n(\omega) < \infty\}) = 1$ .*

As a consequence of the above theorem, we have the following.

- (a) Let  $\omega \in \{\omega \in \Omega : \sum_{n=1}^{\infty} \chi_n(\omega) < \infty\}$ . Then there exists an  $N \geq 1$  and  $R > 0$  (depending on  $\omega$ ) such that, for every  $n \geq N$ ,  $X_n(\omega) = X'_n(\omega)$  and  $\sup_{n \geq N} \|X_n(\omega)\| \leq R$ . Therefore  $\sum_{n \geq N} \mathbb{E}[(a(n))^2 \|M_{n+1}\|^2 | \mathcal{F}_n](\omega) \leq \sum_{n \geq N} (a(n))^2 K^2 (1 + \|X'_n(\omega)\|)^2 \leq K^2 (1 + R)^2 \sum_{n \geq N} (a(n))^2 < \infty$ , where the last inequality follows from assumption (A2)(ii). Therefore,

$$\{\omega \in \Omega : \sum_{n \geq 1} \chi_n(\omega) < \infty\} \subseteq \{\omega \in \Omega : \sum_{n=0}^{\infty} \mathbb{E}[(a(n))^2 \|M_{n+1}\|^2 | \mathcal{F}_n](\omega) < \infty\}.$$

Therefore by Theorem 4.1, we have that the  $\mathbb{P}(\sum_{n=0}^{\infty} \mathbb{E}[(a(n))^2 \|M_{n+1}\|^2 | \mathcal{F}_n] < \infty) = 1$  and by martingale convergence theorem (see [3, Section 11.3, Thm. 11]) we have that, the square integrable martingale  $\{\sum_{m=0}^{n-1} a(m)M_{m+1}, \mathcal{F}_n\}_{n \geq 1}$  converges almost surely.

- (b) Thus for  $\omega$  lying in a probability one set, there exists  $N \geq 1$  and  $R > 0$  (depending on  $\omega$ ) such that along this sample path the iterates  $\{X_n(\omega)\}_{n \geq N}$ , can be viewed as being generated by recursion (5) with initial condition  $X_N(\omega)$ , their norms are bounded by  $R$  uniformly and the additive noise terms  $\{M_n(\omega)\}_{n \geq N}$  satisfy the hypothesis of [2, Prop. 1.3]. Then by arguments similar to those of Theorem 3.1(b) we have that,

**Corollary 4.2** *Under assumptions (A1) – (A3) and (A4)', for almost every  $\omega$ , the iterates generated by Algorithm 1,  $\{X'_n(\omega)\}_{n \geq 0}$ , are such that  $X'_n(\omega) \rightarrow A$  as  $n \rightarrow \infty$ .*

## 5 Proof of the lock-in probability theorem (Thm. 3.2)

Proof of the lock-in probability result follows as a consequence of a series of lemmas. The overall structure can be summarized as follows.

- (a) Our first aim is to replace the set-valued map in recursion (5) with an equivalent single-valued locally Lipschitz continuous function with an additional parameter. In order to accomplish this, we first embed the graph of the set-valued map  $F$  in the graph of a sequence of locally Lipschitz continuous set-valued maps. These maps are then parametrized using the Stiener selection procedure which preserves the modulus of continuity.
- (b) The relation between the solutions of DI (6) and that of differential inclusions with continuous set-valued maps which approximate  $F$  (as in (a) above) is established.
- (c) An ordinary differential equation(o.d.e) is defined using an appropriate single-valued parametrization of  $F$  (as in (a) above). The existence of solutions to such an o.d.e. and further its uniqueness follow from Caratheodary's existence theorem and locally Lipschitz nature of the vector field respectively. The solutions of this o.d.e. aide in separating the probability contributions due to the additive noise terms and the set-valued nature of the drift function. Using the results from part (b) above, we conclude that after a large number of iterations, the probability contribution is only due to the additive noise terms.

- (d) We finally review the standard probability lower bounding procedure for the additive noise terms from [3, Ch. 4]. Using this bound in the result obtained in part (c) above gives us the desired lock-in probability bound.

Throughout, we use  $U$  to denote the closed unit ball in  $\mathbb{R}^d$  centered at the origin. Further, for every  $Y_1, Y_2 \subseteq \mathbb{R}^d$  and  $r \in \mathbb{R}$ , define,

- $Y_1 + Y_2 := \{y_1 + y_2 : y_1 \in Y_1 \text{ and } y_2 \in Y_2\}$ ,
- $rY_1 := \{ry_1 : y_1 \in Y_1\}$ .

## 5.1 Upper semicontinuous set-valued maps and their approximation

First we recall definitions of continuous set-valued maps and locally Lipschitz continuous set-valued maps. These notions are taken from [12, Ch. 1].

**Definition** A set-valued map  $F : \mathbb{R}^d \rightarrow \{\text{compact subsets of } \mathbb{R}^d\}$  is,

- *upper semicontinuous* (u.s.c.) if, for every  $x \in \mathbb{R}^d$ , for every  $\epsilon > 0$ , there exists a  $\delta > 0$  (depending on  $x$  and  $\epsilon$ ) such that, for every  $x' \in \mathbb{R}^d$  satisfying  $\|x' - x\| < \delta$ , we have that  $F(x') \subseteq F(x) + \epsilon U$ , where  $F(x) + \epsilon U := \{y + \epsilon u : y \in F(x), u \in U\}$ .
- *lower semicontinuous* (l.s.c.) if, for every  $x \in \mathbb{R}^d$ , for every  $\mathbb{R}^d$ -valued sequence  $\{x_n\}_{n \geq 1}$  converging to  $x$ , for every  $y \in F(x)$ , there exists a sequence  $\{y_n \in F(x_n)\}_{n \geq 1}$  converging to  $y$ .
- *continuous* if, it is both u.s.c. and l.s.c.
- *locally Lipschitz continuous* if, for every  $x_0 \in \mathbb{R}^d$ , there exists  $\delta > 0$  and  $L > 0$  (depending on  $x_0$ ) such that for every  $x, x' \in x_0 + \delta U$ , we have that  $F(x) \subseteq F(x') + L\|x - x'\|U$ .

Let  $\mathcal{K}(\mathbb{R}^d)$  denote the family of all non-empty compact subsets of  $\mathbb{R}^d$ . Let  $\mathbf{H} : \mathcal{K}(\mathbb{R}^d) \times \mathcal{K}(\mathbb{R}^d) \rightarrow [0, \infty)$  be defined such that, for every  $S_1, S_2 \in \mathcal{K}(\mathbb{R}^d)$ ,

$$\mathbf{H}(S_1, S_2) := \max \left\{ \sup_{s_1 \in S_1} \inf_{s_2 \in S_2} \|s_1 - s_2\|, \sup_{s_2 \in S_2} \inf_{s_1 \in S_1} \|s_1 - s_2\| \right\}. \quad (13)$$

With  $\mathbf{H}$  as defined above,  $(\mathcal{K}(\mathbb{R}^d), \mathbf{H})$  is a complete metric space (for a proof see [13, Thm. 1.1.2]). The notions of continuity and local Lipschitz continuity of a set-valued map can be restated using the metric defined above and is stated as a lemma below for easy reference (for a proof see [12, Ch. 1, section 5, Cor. 1]).

**Lemma 5.1** A set-valued map  $F : \mathbb{R}^d \rightarrow \mathcal{K}(\mathbb{R}^d)$  is

- (a) *Continuous*, if and only if, for every  $x_0 \in \mathbb{R}^d$ , for every  $\epsilon > 0$ , there exists  $\delta > 0$  (depending on  $x_0$  and  $\epsilon$ ), such that for every  $x \in x_0 + \delta U$ ,  $\mathbf{H}(F(x), F(x_0)) < \epsilon$ .
- (b) *locally Lipschitz continuous*, if and only if, for every  $x_0 \in \mathbb{R}^d$ , there exists  $\delta > 0$  and  $L > 0$  (depending on  $x_0$ ), such that for every  $x, x' \in x_0 + \delta U$ ,  $\mathbf{H}(F(x), F(x')) \leq L\|x - x'\|$ .

Before we proceed further we look at a certain form of locally Lipschitz continuous set-valued maps that arise later. The next lemma defines such maps and also states that the sum of two locally Lipschitz continuous set-valued maps is again a locally Lipschitz continuous set-valued map, a result needed later to obtain locally Lipschitz continuous single-valued parametrization of map  $F$  in recursion (5).

**Lemma 5.2** (a) If  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is a locally Lipschitz continuous map and  $C \in \mathcal{K}(\mathbb{R}^d)$ , then the set-valued map  $F : \mathbb{R}^d \rightarrow \mathcal{K}(\mathbb{R}^d)$ , given by  $F(x) := f(x)C$  for every  $x \in \mathbb{R}^d$ , is a locally Lipschitz continuous set-valued map.

- (b) If for every  $i \in \{1, 2\}$ ,  $F_i : \mathbb{R}^d \rightarrow \mathcal{K}(\mathbb{R}^d)$  is a locally Lipschitz continuous set-valued map, then the set-valued map  $F : \mathbb{R}^d \rightarrow \mathcal{K}(\mathbb{R}^d)$ , given by  $F(x) := F_1(x) + F_2(x)$  for every  $x \in \mathbb{R}^d$ , is a locally Lipschitz continuous set-valued map.

**Proof** (a) Fix  $x_0 \in \mathbb{R}^d$  and let  $r := \sup_{c \in C} \|c\|$ . Since  $f$  is locally Lipschitz continuous, there exists  $\delta_{x_0}^f > 0$  and  $L_{x_0}^f > 0$  such that for every  $x, x' \in x_0 + \delta_{x_0}^f U$ ,  $|f(x) - f(x')| \leq L_{x_0}^f \|x - x'\|$ . Let  $x, x' \in x_0 + \delta_{x_0}^f U$ . Then for any  $c \in C$ ,

$$\begin{aligned} \|f(x)c - f(x')c\| &= |f(x) - f(x')| \|c\| \\ &\leq r L_{x_0}^f \|x - x'\|. \end{aligned}$$

Therefore for every  $x, x' \in x_0 + \delta_{x_0}^f U$ , for every  $c \in C$ ,  $f(x')c - f(x)c \in r L_{x_0}^f \|x - x'\| U$ . Thus for every  $x, x' \in x_0 + \delta_{x_0}^f U$ ,  $F(x') \subseteq F(x) + r L_{x_0}^f \|x - x'\| U$ , from which it follows that the set-valued map  $F$  is locally Lipschitz continuous at  $x_0$  with  $\delta := \delta_{x_0}^f$  and  $L := r L_{x_0}^f$ . Since  $x_0 \in \mathbb{R}^d$  is arbitrary, the above argument gives us that  $F$  is locally Lipschitz continuous at every  $x_0$ .

(b) Fix  $x_0 \in \mathbb{R}^d$ . Since for every  $i \in \{1, 2\}$ ,  $F_i$  are locally Lipschitz continuous, there exists  $\delta_i > 0$  and  $L_i > 0$  such that for every  $x, x' \in x_0 + \delta_i U$ ,  $F_i(x) \subseteq F_i(x') + L_i \|x - x'\| U$ . Let  $\delta := \min\{\delta_1, \delta_2\}$ ,  $L := L_1 + L_2$  and  $x, x' \in x_0 + \delta U$ . For any  $y \in F(x)$ , there exists  $y_1 \in F_1(x)$  and  $y_2 \in F_2(x)$  such that  $y = y_1 + y_2$ . By our choice of  $\delta$ , we have  $y'_1 \in F_1(x')$ ,  $y'_2 \in F_2(x')$  and  $u_1, u_2 \in U$ , such that for every  $i \in \{1, 2\}$ ,  $y_i = y'_i + L_i \|x - x'\| u_i$ . Therefore,

$$\begin{aligned} y &= y_1 + y_2 \\ &= y'_1 + L_1 \|x - x'\| u_1 + y'_2 + L_2 \|x - x'\| u_2 \\ &= y'_1 + y'_2 + (L_1 + L_2) \|x - x'\| \left( \frac{L_1 u_1 + L_2 u_2}{L_1 + L_2} \right). \end{aligned} \tag{14}$$

Clearly  $y'_1 + y'_2 \in F(x')$  and since  $U$  is a convex subset of  $\mathbb{R}^d$ ,  $\frac{L_1 u_1 + L_2 u_2}{L_1 + L_2} \in U$ . From (14) we get that,  $F(x) \subseteq F(x') + (L_1 + L_2) \|x - x'\| U$ , for every  $x, x' \in x_0 + \delta U$ . Therefore  $F$  is locally Lipschitz continuous at  $x_0$ . Since  $x_0$  is arbitrary, the above argument gives us that  $F$  is locally Lipschitz continuous. ■

Consider a set-valued map  $F$  satisfying assumption (A1). A simple contradiction argument gives us that  $F$  is u.s.c. It is not possible to represent such u.s.c. set-valued maps with a single-valued continuous map with an additional parameter. But instead one can approximate them from above as explained next. The first step is to embed the graph of the map  $F$  in that of a sequence of continuous set-valued maps as stated in the lemma below. For the proof of the lemma below notions of a paracompact topological space, an open covering, its locally finite refinement and partition of unity subordinated to a locally finite covering are needed, which are summarized in Appendix A for easy reference.

**Lemma 5.3** *Let  $F : \mathbb{R}^d \rightarrow \mathcal{K}(\mathbb{R}^d)$  be a set-valued map satisfying (A1). Then, there exists a sequence of continuous set-valued maps  $\{F^{(l)} : \mathbb{R}^d \rightarrow \mathcal{K}(\mathbb{R}^d)\}_{l \geq 1}$ , such that for every  $l \geq 1$ ,*

- (a) *for every  $x \in \mathbb{R}^d$ ,  $F^{(l)}(x)$  is a non-empty, convex and compact subset of  $\mathbb{R}^d$ ,*
- (b) *for every  $x \in \mathbb{R}^d$ ,  $F(x) \subseteq F^{(l+1)}(x) \subseteq F^{(l)}(x)$ ,*
- (c) *there exists  $K^{(l)} > 0$ , such that for every  $x \in \mathbb{R}^d$ ,  $\sup_{y \in F^{(l)}(x)} \|y\| \leq K^{(l)}(1 + \|x\|)$ ,*
- (d)  *$F^{(l)}$  is a locally Lipschitz continuous set valued map.*

Furthermore,

- (e) *for every  $x \in \mathbb{R}^d$ ,  $F(x) = \cap_{l \geq 1} F^{(l)}(x)$ .*

**Proof** For any  $\epsilon > 0$ , for every  $x_0 \in \mathbb{R}^d$ , let  $B(\epsilon, x_0) := \{x : \|x - x_0\| < \epsilon\}$ . Let  $\{\epsilon_l := \frac{1}{3^l}\}_{l \geq 1}$ . Then for every  $l \geq 1$ ,  $\mathcal{G}_l := \{B(\epsilon_l, x_0) : x_0 \in \mathbb{R}^d\}$  is an open covering of  $\mathbb{R}^d$ . Since  $\mathbb{R}^d$  is a metric space, it is paracompact (see [12, Ch. 0, Sec. 1, Thm. 1]). Therefore for every  $l \geq 1$ , there exists a locally finite open refinement of the covering  $\mathcal{G}_l$  and let it be denoted by  $\tilde{\mathcal{G}}_l := \{C_i^l\}_{i \in I^l}$  where  $I^l$  is an arbitrary index set. By [12, Ch. 0, Sec. 1, Thm. 2], there exists a locally Lipschitz continuous partition of unity,  $\{\psi_i^l\}_{i \in I^l}$ , subordinated to the covering  $\tilde{\mathcal{G}}_l$ . Therefore, for every  $l \geq 1$ , for every  $i \in I^l$ , there exists  $x_i^l$ , such that  $\text{support}(\psi_i^l) \subseteq C_i^l \subseteq B(\epsilon_l, x_i^l)$ . For every  $l \geq 1$ , for every  $x \in \mathbb{R}^d$ , let  $I^l(x) := \{i \in I^l : \psi_i^l(x) > 0\}$  and by definition of  $\psi_i^l$ , we have that  $0 < |I^l(x)| < \infty$  and  $\sum_{i \in I^l(x)} \psi_i^l(x) = 1$ .

For every  $l \geq 1$ , define the set valued map  $F^{(l)} : \mathbb{R}^d \rightarrow \{\text{subsets of } \mathbb{R}^d\}$ , such that for every  $x \in \mathbb{R}^d$ ,  $F^{(l)}(x) := \sum_{i \in I^l(x)} \psi_i^l(x) A_i^l$ , where  $A_i^l := \text{co}(F(B(2\epsilon_l, x_i^l)))$ .

The proofs of parts (a), (b), (c) and (e) of the lemma are exactly the same as that of [14, Lemma 3.2]. We shall provide a proof of part (d) of the lemma above from which continuity of the set-valued maps  $F^{(l)}$  follows.

- (d) Fix  $l \geq 1$  and  $x \in \mathbb{R}^d$ . Since  $\tilde{\mathcal{C}}_l$  is a locally finite open covering of  $\mathbb{R}^d$ , there exists  $\delta > 0$  (depending on  $x$ ), such that  $I^l(x, \delta) := \{i \in I^l : B(x, \delta) \cap C_i^l \neq \emptyset\}$  is finite. Since  $\{\psi_i^l\}_{i \in I^l}$  is a locally Lipschitz continuous partition of unity subordinated to the covering  $\tilde{\mathcal{C}}_l$ , we have that for every  $i \in I^l$ ,  $\text{support}(\psi_i^l) \subseteq C_i^l$ . Therefore, for every  $x' \in B(x, \delta)$ ,  $F^{(l)}(x') = \sum_{i \in I^l(x, \delta)} \psi_i^l(x') A_i^l$ .

From the proof of part (a) of this lemma we know that for every  $i \in I^l$ ,  $A_i^l$  is a compact and convex subset of  $\mathbb{R}^d$ . Therefore from Lemma 5.2(a), we get that, for every  $i \in I^l(x, \delta)$ , the set-valued map given by  $y \rightarrow \psi_i^l(y) A_i^l$  is locally Lipschitz continuous. Further since  $|I^l(x, \delta)| < \infty$ , from Lemma 5.2(b), we get that the set-valued map given by  $y \rightarrow \sum_{i \in I^l(x, \delta)} \psi_i^l(y) A_i^l$  is locally Lipschitz continuous. Since the set-valued map  $y \rightarrow \sum_{i \in I^l(x, \delta)} \psi_i^l(y) A_i^l$  restricted to  $B(x, \delta)$  is the same as  $F^{(l)}$  on  $B(x, \delta)$ , we get that  $F^{(l)}$  is locally Lipschitz continuous at  $x$ . Since  $x$  is arbitrary, the above argument gives us that  $F^{(l)}$  is a locally Lipschitz continuous set-valued map. ■

The continuous set-valued maps  $F^{(l)}$  as obtained above can be now parametrized (that is represented with a single-valued continuous function with an additional parameter). Key to parametrization is a continuous selection procedure by which we mean a function  $\sigma : \mathcal{K}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$  which is continuous and is such that for every  $Y \in \mathcal{K}(\mathbb{R}^d)$ ,  $\sigma(Y) \in Y$ . Since the maps  $F^{(l)}$  are convex set-valued, it suffices to look for a selection procedure which is continuous restricted to the family of compact and convex subsets of  $\mathbb{R}^d$ . Further we want a selection procedure which would preserve the local Lipschitz continuity of the set-valued map  $F^{(l)}$  in the parametrization as well. In order to accomplish this we shall use the Stienier selection procedure (for a definition see [15, Thm. 9.4.1]). The next lemma summarizes some properties of the Stienier selection procedure and an intersection lemma which form the central tools for parameterizing the set-valued maps  $F^{(l)}$  (for a proof we refer the reader to [15, Thm. 9.4.1] and [15, Lemma 9.4.2]). Before we state the lemma we introduce some notation needed. Let  $\mathcal{K}_c(\mathbb{R}^d)$  denote the family of all non-empty compact and convex subsets of  $\mathbb{R}^d$ . For any set  $Y \subseteq \mathbb{R}^d$  and for any  $x \in \mathbb{R}^d$ , define  $\mathbf{d}(x, Y) := \inf_{y \in Y} \|x - y\|$ .

**Lemma 5.4** (a) *There exists a function  $\sigma : \mathcal{K}_c(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ , such that for every  $Y, Y_1, Y_2 \in \mathcal{K}_c(\mathbb{R}^d)$ ,*

$$\sigma(Y) \in Y \text{ and } \|\sigma(Y_1) - \sigma(Y_2)\| \leq d \mathbf{H}(Y_1, Y_2).$$

- (b) *The map  $\Pi : \mathcal{K}_c(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathcal{K}_c(\mathbb{R}^d)$ , defined such that for every  $Y \in \mathcal{K}_c(\mathbb{R}^d)$  and  $x \in \mathbb{R}^d$ ,  $\Pi(Y, x) := Y \cap (x + 2\mathbf{d}(x, Y)U)$ , is such that for every  $Y_1, Y_2 \in \mathcal{K}_c(\mathbb{R}^d)$  and for every  $x_1, x_2 \in \mathbb{R}^d$ ,*

$$\mathbf{H}(\Pi(Y_1, x_1), \Pi(Y_2, x_2)) \leq 5 (\mathbf{H}(Y_1, Y_2) + \|x_1 - x_2\|).$$

We now use the results stated in the above lemma to parametrize the set-valued maps  $F^{(l)}$ .

**Lemma 5.5** *Let  $\{F^{(l)}\}_{l \geq 1}$  be as in Lemma 5.3. For every  $l \geq 1$ , there exists a continuous function  $f^{(l)} : \mathbb{R}^d \times U \rightarrow \mathbb{R}^d$  such that,*

- (a) *for every  $x \in \mathbb{R}^d$ ,  $f^{(l)}(x, U) = F^{(l)}(x)$  where  $f^{(l)}(x, U) := \{f^{(l)}(x, u) : u \in U\}$ .*  
(b) *for  $K^{(l)} > 0$  as in Lemma 5.3, for every  $(x, u) \in \mathbb{R}^d \times U$ ,  $\|f^{(l)}(x, u)\| \leq K^{(l)}(1 + \|x\|)$ .*  
(c) *for every  $x_0 \in \mathbb{R}^d$ , there exists  $\delta^{(l)} > 0$  and  $L^{(l)} > 0$  (depending on  $x_0$ ), such that for every  $x, x' \in x_0 + \delta^{(l)}U$ , for every  $u \in U$ ,*

$$\|f^{(l)}(x, u) - f^{(l)}(x', u)\| \leq L^{(l)} \|x - x'\|.$$

**Proof** Fix  $l \geq 1$ . Let the map  $f^{(l)} : \mathbb{R}^d \times U \rightarrow \mathbb{R}^d$  be defined such that, for every  $(x, u) \in \mathbb{R}^d \times U$ ,

$$f^{(l)}(x, u) := \sigma \left( \Pi \left( F^{(l)}(x), K^{(l)}(1 + \|x\|)u \right) \right), \quad (15)$$

where  $\sigma$  and  $\Pi$  are as in Lemma 5.4.

(a) By definition of  $f^{(l)}$ ,  $\sigma$  and  $\Pi$ , for every  $(x, u) \in \mathbb{R}^d \times U$ , we have that,

$$f^{(l)}(x, u) \in \Pi(F^{(l)}(x), K^{(l)}(1 + \|x\|)u) \subseteq F^{(l)}(x).$$

Therefore, for every  $x \in \mathbb{R}^d$ ,  $f^{(l)}(x, U) \subseteq F^{(l)}(x)$ . By Lemma 5.3(c), we know that for every  $x \in \mathbb{R}^d$ ,  $\sup_{y \in F^{(l)}(x)} \|y\| \leq K^{(l)}(1 + \|x\|)$ . Thus for every  $x \in \mathbb{R}^d$ , for any  $y \in F^{(l)}(x)$ , there exists  $u \in U$ , such that  $y = K^{(l)}(1 + \|x\|)u$ . For such a  $u \in U$ , by definition of  $\Pi$ , we have that  $\Pi(F^{(l)}(x), K^{(l)}(1 + \|x\|)u) = y$  and hence  $f^{(l)}(x, u) = \sigma(\Pi(F^{(l)}(x), K^{(l)}(1 + \|x\|)u)) = y$ . Therefore for every  $x \in \mathbb{R}^d$ ,  $F^{(l)}(x) \subseteq f^{(l)}(x, U)$  from which it follows that  $f^{(l)}(x, U) = F^{(l)}(x)$ , for every  $x \in \mathbb{R}^d$ .

(b) Follows from part (a) of this lemma and Lemma 5.3(c).

(c) Fix  $x_0 \in \mathbb{R}^d$ . Since  $F^{(l)}$  is a locally Lipschitz continuous set-valued map (see Lemma 5.3(d)), we obtain  $\delta_{F^{(l)}} > 0$  and  $L_{F^{(l)}} > 0$  (depending on  $x_0$ ) such that for every  $x, x' \in x_0 + \delta_{F^{(l)}}U$ ,  $\mathbf{H}(F^{(l)}(x), F^{(l)}(x')) \leq L_{F^{(l)}}\|x - x'\|$ . Set  $\delta^{(l)} := \delta_{F^{(l)}}$  and  $L^{(l)} := 5d(L_{F^{(l)}} + K^{(l)})$ . Then, for any  $x, x' \in x_0 + \delta^{(l)}U$ , for every  $u \in U$ ,

$$\begin{aligned} \|f^{(l)}(x, u) - f^{(l)}(x', u)\| &= \left\| \sigma\left(\Pi\left(F^{(l)}(x), K^{(l)}(1 + \|x\|)u\right)\right) - \sigma\left(\Pi\left(F^{(l)}(x'), K^{(l)}(1 + \|x'\|)u\right)\right) \right\| \\ &\leq d \mathbf{H}\left(\Pi\left(F^{(l)}(x), K^{(l)}(1 + \|x\|)u\right), \Pi\left(F^{(l)}(x'), K^{(l)}(1 + \|x'\|)u\right)\right) \end{aligned} \quad (16)$$

$$\leq 5d \left( \mathbf{H}\left(F^{(l)}(x), F^{(l)}(x')\right) + \left\| K^{(l)}(1 + \|x\|)u - K^{(l)}(1 + \|x'\|)u \right\| \right) \quad (17)$$

$$\begin{aligned} &= 5d \left( \mathbf{H}\left(F^{(l)}(x), F^{(l)}(x')\right) + K^{(l)}\| \|x\| - \|x'\| \| \|u\| \right) \\ &\leq 5d \left( \mathbf{H}\left(F^{(l)}(x), F^{(l)}(x')\right) + K^{(l)}\|x - x'\| \right) \\ &\leq 5d \left( L_{F^{(l)}}\|x - x'\| + K^{(l)}\|x - x'\| \right) \\ &= L^{(l)}\|x - x'\|, \end{aligned} \quad (18)$$

where, (16) follows from Lemma 5.4(a), (17) follows from Lemma 5.4(b) and (18) follows from our choice of  $\delta^{(l)}$  and local Lipschitz continuity of  $F^{(l)}$ . ■

The set-valued map in recursion (5) can be replaced with the parametrization obtained in the lemma above as explained below.

(1) For every  $l \geq 1$ , by Lemma 5.3(b), we know that for every  $x \in \mathbb{R}^d$ ,  $F(x) \subseteq F^{(l)}(x)$ . Therefore for every  $l \geq 1$ , for every  $n \geq 0$ ,

$$X_{n+1} - X_n - a(n)M_{n+1} \in a(n)F^{(l)}(X_n).$$

(2) For every  $l \geq 1$ , by Lemma 5.5(a), we know that for every  $x \in \mathbb{R}^d$ ,  $F^{(l)}(x) = f^{(l)}(x, U)$ . It can now be shown that for every  $n \geq 0$ , there exists a  $U$ -valued random variable on  $\Omega$ , say  $U_n^{(l)}$ , such that for every  $\omega \in \Omega$ , for every  $n \geq 0$ ,

$$X_{n+1}(\omega) - X_n(\omega) - a(n)M_{n+1}(\omega) = a(n)f^{(l)}(X_n(\omega), U_n^{(l)}(\omega)) \quad (19)$$

(for a proof see [14, Lemma 6.1]).

## 5.2 Solutions of the mean field and their approximation

In this section, we shall approximate the solutions of mean field (that is DI (6)) with the solutions of DI given by,

$$\frac{dx}{dt} \in F^{(l)}(x), \quad (20)$$

for some  $l \geq 1$ . In order to accomplish this we need some notations which are introduced next.

For every  $T > 0$  and for every  $x \in \mathbb{R}^d$ , let  $S(T, x)$  denote the set of solutions of DI (6) on  $[0, T]$ . Formally,

$$S(T, x) := \left\{ \mathbf{x} : [0, T] \rightarrow \mathbb{R}^d : \mathbf{x} \text{ is absolutely continuous with } \mathbf{x}(0) = x \text{ and for a.e. } t \in [0, T], \frac{d\mathbf{x}(t)}{dt} \in F(\mathbf{x}(t)) \right\}. \quad (21)$$

Since  $F$  is a Marchaud map, we have that for every  $T > 0$  and for every  $x \in \mathbb{R}^d$ ,  $S(T, x) \neq \emptyset$ . Similarly for every  $l \geq 1$ , for every  $T > 0$  and for every  $x \in \mathbb{R}^d$ , let  $S^{(l)}(T, x)$  denote the set of solutions of DI (20) on  $[0, T]$ . Formally,

$$S^{(l)}(T, x) := \left\{ \mathbf{x} : [0, T] \rightarrow \mathbb{R}^d : \mathbf{x} \text{ is absolutely continuous with } \mathbf{x}(0) = x \text{ and for a.e. } t \in [0, T], \frac{d\mathbf{x}(t)}{dt} \in F^{(l)}(\mathbf{x}(t)) \right\}. \quad (22)$$

From Lemma 5.3, we know that for every  $l \geq 1$ ,  $F^{(l)}$  is a Marchaud map and hence for every  $T > 0$  and for every  $x \in \mathbb{R}^d$ ,  $S^{(l)}(T, x) \neq \emptyset$ .

For any  $Y \subseteq \mathbb{R}^d$ , for any  $T > 0$ , define  $S(T, Y) := \cup_{y \in Y} S(T, y)$ . Similarly, for every  $l \geq 1$ ,  $S^{(l)}(T, Y) := \cup_{y \in Y} S^{(l)}(T, y)$ .

The next lemma summarizes some important relationships between the solutions of DI (6) and those of DI (20) needed later. It also states that for large enough  $l \geq 1$ , the solutions of DI (20) are within an  $\epsilon$ -neighborhood of the solutions of DI (6) for every initial condition lying in a compact subset of  $\mathbb{R}^d$ .

**Lemma 5.6** *For every  $T > 0$ ,*

- (a) *for every  $l \geq 1$ , for every  $x \in \mathbb{R}^d$ ,  $S(T, x) \subseteq S^{(l+1)}(T, x) \subseteq S^{(l)}(T, x)$ .*
- (b) *for every  $x \in \mathbb{R}^d$ ,  $S(T, x) = \cap_{l \geq 1} S^{(l)}(T, x)$ .*
- (c) *for any  $Y \subseteq \mathbb{R}^d$ ,  $S(T, Y) = \cap_{l \geq 1} S^{(l)}(T, Y)$ .*
- (d) *for every  $Y \subseteq \mathbb{R}^d$  compact,  $S(T, Y)$  is a compact subset of  $\mathcal{C}([0, T], \mathbb{R}^d)$  (the vector space of  $\mathbb{R}^d$ -valued continuous functions on  $[0, T]$ ).*
- (e) *for every  $Y \subseteq \mathbb{R}^d$  compact, for every  $l \geq 1$ ,  $S^{(l)}(T, Y)$  is a compact subset of  $\mathcal{C}([0, T], \mathbb{R}^d)$ .*
- (f) *for every  $Y \subseteq \mathbb{R}^d$  compact, for every  $\epsilon > 0$ , there exists  $l' \geq 1$ , such that for every  $l \geq l'$ , for every  $\mathbf{x}^{(l)} \in S^{(l)}(T, Y)$ , there exists  $\mathbf{x} \in S(T, Y)$ , such that  $\sup_{t \in [0, T]} \|\mathbf{x}(t) - \mathbf{x}^{(l)}(t)\| < \epsilon$ .*

**Proof** Fix  $T > 0$ .

(a) Fix  $l \geq 1$  and  $x \in \mathbb{R}^d$ . Let  $\mathbf{x} \in S(T, x)$ . Then we have that  $\mathbf{x}$  is absolutely continuous with  $\mathbf{x}(0) = x$  and for a.e.  $t \in [0, T]$ ,  $\frac{d\mathbf{x}(t)}{dt} \in F(\mathbf{x}(t))$ . By Lemma 5.3(b), we know that for every  $t \in [0, T]$ ,  $F(\mathbf{x}(t)) \subseteq F^{(l+1)}(\mathbf{x}(t))$ . Therefore for a.e.  $t \in [0, T]$ ,  $\frac{d\mathbf{x}(t)}{dt} \in F^{(l+1)}(\mathbf{x}(t))$ , from which we get that  $\mathbf{x} \in S^{(l+1)}(T, x)$ . Hence  $S(T, x) \subseteq S^{(l+1)}(T, x)$ . Using the fact that for every  $x' \in \mathbb{R}^d$ ,  $F^{(l+1)}(x') \subseteq F^{(l)}(x')$  (see Lemma 5.3(b)), a similar argument gives us that  $S^{(l+1)}(T, x) \subseteq S^{(l)}(T, x)$ .

(b) Fix  $x \in \mathbb{R}^d$ . From part (a) of this lemma we have that  $S(T, x) \subseteq \cap_{l \geq 1} S^{(l)}(T, x)$ . Let  $\mathbf{x} \in \cap_{l \geq 1} S^{(l)}(T, x)$ . Then  $\mathbf{x}$  is absolutely continuous with  $\mathbf{x}(0) = x$  and for every  $l \geq 1$ , for a.e.  $t \in [0, T]$ ,  $\frac{d\mathbf{x}(t)}{dt} \in F^{(l)}(\mathbf{x}(t))$ . Thus for a.e.  $t \in [0, T]$ , for every  $l \geq 1$ ,  $\frac{d\mathbf{x}(t)}{dt} \in F^{(l)}(\mathbf{x}(t))$ . Hence for a.e.  $t \in [0, T]$ ,  $\frac{d\mathbf{x}(t)}{dt} \in \cap_{l \geq 1} F^{(l)}(\mathbf{x}(t)) = F(\mathbf{x}(t))$ , where the equality follows from Lemma 5.3(e). Therefore  $\mathbf{x} \in S(T, x)$ , from which we get that  $\cap_{l \geq 1} S^{(l)}(T, x) \subseteq S(T, x)$ .

(c) Follows from part (a) and (b) of this lemma.

(d) & (e) Follows from [2, Lemma 3.1].

(f) Suppose not. Then there exists  $Y \subseteq \mathbb{R}^d$  compact and  $\epsilon > 0$ , such that for every  $l' \geq 1$ , there exists  $l \geq l'$  and  $\mathbf{x}^{(l)} \in S^{(l)}(T, Y)$ , such that  $d(\mathbf{x}^{(l)}, S(T, Y)) \geq \epsilon$ , where  $d(\mathbf{x}^{(l)}, S(T, Y)) := \inf_{\mathbf{x} \in S(T, Y)} \sup_{t \in [0, T]} \|\mathbf{x}^{(l)}(t) - \mathbf{x}(t)\|$ . Thus we can obtain a sequence of solutions, say  $\{\mathbf{x}^{(l_k)}\}_{k \geq 1}$ , such that for every  $k \geq 1$ ,  $1 \leq l_k < l_{k+1}$  and  $\mathbf{x}^{(l_k)} \in S^{(l_k)}(T, Y)$  with  $d(\mathbf{x}^{(l_k)}, S(T, Y)) \geq \epsilon$ . From part (a) of this lemma, we have that for every  $k \geq 1$ ,  $S^{(l_k)}(T, Y) \subseteq S^{(1)}(T, Y)$  and hence  $\{\mathbf{x}^{(l_k)}\}_{k \geq 1} \subseteq S^{(1)}(T, Y)$ . Since  $Y \subseteq \mathbb{R}^d$  is compact, by part (e) of this lemma we know that  $S^{(1)}(T, Y)$  is a compact subset of  $\mathcal{C}([0, T], \mathbb{R}^d)$ . Thus there exists a subsequence of  $\{\mathbf{x}^{(l_k)}\}_{k \geq 1}$ , say  $\{\mathbf{x}^{(l_{k_j})}\}_{j \geq 1}$  such that  $\mathbf{x}^{(l_{k_j})} \rightarrow \mathbf{x}^*$  as  $j \rightarrow \infty$  in  $\mathcal{C}([0, T], \mathbb{R}^d)$  and  $\mathbf{x}^* \in S^{(1)}(T, Y)$ . Since for every

$j \geq 1$ ,  $d(\mathbf{x}^{(l_{k_j})}, S(T, Y)) \geq \epsilon$ , we get that  $d(\mathbf{x}^*, S(T, Y)) \geq \epsilon$  and hence  $\mathbf{x}^* \notin S(T, Y)$ . From part (a) of this lemma, we get that for every  $l \geq 1$ , for  $J := \min\{j \geq 1 : l_{k_j} \geq l\}$ ,  $\{\mathbf{x}^{(l_{k_j})}\}_{j \geq J} \subseteq S^{(l)}(T, Y)$ . Further by part (e) of this lemma we have that for every  $l \geq 1$ ,  $S^{(l)}(T, Y)$  is a compact subset of  $\mathcal{C}([0, T], \mathbb{R}^d)$ . Thus for every  $l \geq 1$ ,  $\mathbf{x}^* \in S^{(l)}(T, Y)$  and hence  $\mathbf{x}^* \in \cap_{l \geq 1} S^{(l)}(T, Y) = S(T, Y)$  (see part (c) of this lemma). This leads to a contradiction. ■

The part (f) of the above lemma provides the necessary approximation result. Further since the set-valued maps  $F^{(l)}$  admit a single-valued parametrization ( $f^{(l)}$  as in Lemma 5.5), a solution of DI (20) can be viewed as a solution of the ordinary differential equation (o.d.e.) given by,

$$\frac{dx}{dt} = f^{(l)}(x, u(t)), \quad (23)$$

for some  $u : [0, \infty) \rightarrow U$  measurable and vice versa. The lemma below summarizes some useful results on the solutions of o.d.e. (23) and its vector field.

**Lemma 5.7** *For every  $l \geq 1$ ,*

- (a) *for every  $T > 0$ , for any  $u : [0, T] \rightarrow U$  measurable, for every initial condition, the set of solutions of o.d.e. (23) is non-empty. That is, for every  $x_0 \in \mathbb{R}^d$ , there exists  $\mathbf{x} : [0, T] \rightarrow \mathbb{R}^d$  such that,  $\mathbf{x}$  is absolutely continuous,  $\mathbf{x}(0) = x_0$  and for a.e.  $t \in [0, T]$ ,  $\frac{d\mathbf{x}(t)}{dt} = f^{(l)}(\mathbf{x}(t), u(t))$ .*
- (b) *for every  $T > 0$ , for every  $Y \subseteq \mathbb{R}^d$  compact, there exists  $C_1(Y, T, l) > 0$ , such that for every  $u : [0, T] \rightarrow U$  measurable, every solution of o.d.e. (23) with initial condition in  $Y$ , say  $\mathbf{x} : [0, T] \rightarrow \mathbb{R}^d$ , satisfies,*

$$\sup_{t \in [0, T]} \|\mathbf{x}(t)\| \leq C_1(Y, T, l).$$

- (c) *for any  $Y \subseteq \mathbb{R}^d$  compact, there exists  $L(Y, l) > 0$ , such that for every  $T > 0$ , for every  $u : [0, T] \rightarrow U$ , the map  $h : Y \times [0, T] \rightarrow \mathbb{R}^d$ , given by  $h(x, t) := f^{(l)}(x, u(t))$  for every  $(x, t) \in Y \times [0, T]$ , satisfies,*

$$\|h(x, t) - h(x', t)\| \leq L(Y, l)\|x - x'\|,$$

*for every  $x, x' \in Y$  and for every  $t \in [0, T]$ .*

- (d) *for every  $T > 0$ , for every  $u : [0, T] \rightarrow U$ , for every initial condition, o.d.e. (23) admits a unique solution.*

**Proof** Fix  $l \geq 1$ .

- (a) Fix  $T > 0$  and  $u : [0, T] \rightarrow U$  measurable. The proof of this part is a direct application of [16, Thm. 3.8]. We show here that the sufficient conditions required to apply the said theorem are satisfied by the vector field of the o.d.e. (23). First, we show that  $f^{(l)}(\cdot, u(\cdot))$  is a Caratheodary function (see [16, Defn. 3.2]). By Lemma 5.5, it is clear that for every  $t \in [0, T]$ , the map  $x \rightarrow f^{(l)}(x, u(t))$  is continuous and for every  $x \in \mathbb{R}^d$ , the map  $t \rightarrow f^{(l)}(x, u(t))$  is measurable. Further by Lemma 5.5(b), we have that for any  $c > 0$ , for every  $x \in \mathbb{R}^d$  with  $\|x\| \leq c$ , for every  $t \in [0, T]$ ,  $\|f^{(l)}(x, t)\| \leq K^{(l)}(1 + c)$ . Thus  $f^{(l)}(\cdot, u(\cdot))$  is a Caratheodary function. Final condition to verify is on the rate of growth of solutions. By Lemma 5.5(b),  $\|f^{(l)}(x, u(t))\| \leq \psi(\|x\|) := K^{(l)}(1 + \|x\|)$ . The function  $\psi : [0, \infty) \rightarrow [0, \infty)$ , is clearly positive everywhere and the function  $\frac{1}{\psi}$  is locally integrable on  $[0, \infty)$ . A simple argument gives us that for every  $r > 0$ , the integral  $\int_r^\infty \frac{d\tilde{r}}{\psi(\tilde{r})}$  can be lower bounded by the tail of  $\frac{1}{K^{(l)}} \sum_{n=1}^\infty \frac{1}{n}$ . Hence for every  $r > 0$ ,  $\int_r^\infty \frac{d\tilde{r}}{\psi(\tilde{r})} = \infty$ . Now [16, Thm. 3.8] can be applied to obtain the required result.
- (b) Fix  $T > 0$  and  $Y \subseteq \mathbb{R}^d$  compact. Since  $Y$  is compact, there exists  $r > 0$  such that  $\sup_{y \in Y} \|y\| \leq r$ . Set  $C_1(Y, T, l) := (r + K^{(l)}T)e^{K^{(l)}T}$ , where  $K^{(l)} > 0$  is as in Lemma 5.5(b). For some  $u : [0, T] \rightarrow U$  measurable and for some  $x_0 \in Y$ , let  $\mathbf{x} : [0, T] \rightarrow \mathbb{R}^d$  be a solution of o.d.e. (23) with initial condition  $x_0$ . Then, for every  $t \in [0, T]$ ,  $\mathbf{x}(t) = x_0 + \int_0^t f^{(l)}(\mathbf{x}(s), u(s))ds$  and hence for every  $t \in [0, T]$

$$\begin{aligned} \|\mathbf{x}(t)\| &\leq \|x_0\| + \int_0^t \|f^{(l)}(\mathbf{x}(s), u(s))\|ds \\ &\leq r + K^{(l)}T + K^{(l)} \int_0^t \|\mathbf{x}(s)\|ds \end{aligned} \quad (24)$$

where, (24) follows from the fact that  $x_0 \in Y$  and Lemma 5.5(b). The required bound follows from (24) and Gronwall's result (see [3, Sec. 11.2.1, Lemma 6]).

- (c) Fix  $Y \subseteq \mathbb{R}^d$  compact. It is enough to show that there exists  $L(Y, l) > 0$ , such that for every  $y_1, y_2 \in Y$ ,  $\sup_{u \in U} \|f^{(l)}(y_1, u) - f^{(l)}(y_2, u)\| \leq L(Y, l)\|y_1 - y_2\|$ . From Lemma 5.5(c), we know that for every  $x_0 \in Y$ , there exists  $\delta(x_0, l) > 0$  and  $L(x_0, l) > 0$ , such that for every  $x, x' \in x_0 + \delta(x_0, l)U$ , for every  $u \in U$ ,  $\|f^{(l)}(y_1, u) - f^{(l)}(y_2, u)\| \leq L(Y, l)\|y_1 - y_2\|$ . Let  $\mathcal{G} := \{x_0 + \frac{\delta(x_0, l)}{2}\mathring{U} : x_0 \in Y\}$ , where,  $\mathring{U}$  denotes the interior of  $U$ . Since  $Y$  is compact and  $\mathcal{G}$  is an open cover of  $Y$ , there exists  $\{x_1, x_2, \dots, x_k\} \subseteq Y$ , such that  $Y \subseteq \cup_{i=1}^k (x_i + \frac{\delta(x_i, l)}{2}\mathring{U})$ . Set  $\delta(Y, l) := \min_{1 \leq i \leq k} \frac{\delta(x_i, l)}{2}$  and  $L_0(Y, l) := \max_{1 \leq i \leq k} L(x_i, l)$ .

Let  $y_1, y_2 \in (Y \times Y) \cap \{(y_1, y_2) : \|y_1 - y_2\| < \delta(Y, l)\}$ . Then we know that there exists  $i \in \{1, \dots, k\}$ , such that  $y_1 \in x_i + \frac{\delta(x_i, l)}{2}\mathring{U}$ . Further since  $\|y_1 - y_2\| < \delta(Y, l) \leq \frac{\delta(x_i, l)}{2}$ , we have that  $y_2 \in x_i + \delta(x_i, l)\mathring{U}$ . Therefore  $y_1, y_2 \in x_i + \delta(x_i, l)U$  and hence, for every  $u \in U$ ,  $\|f^{(l)}(y_1, u) - f^{(l)}(y_2, u)\| \leq L(x_i, l)\|y_1 - y_2\| \leq L_0(Y, l)\|y_1 - y_2\|$ . Thus for every  $y_1, y_2 \in (Y \times Y) \cap \{(y_1, y_2) : \|y_1 - y_2\| < \delta(Y, l)\}$ ,  $\sup_{u \in U} \|f^{(l)}(y_1, u) - f^{(l)}(y_2, u)\| \leq L_0(Y, l)\|y_1 - y_2\|$ .

Let  $E := (Y \times Y) \cap \{(y_1, y_2) : \|y_1 - y_2\| \geq \delta(Y, l)\}$ . By Lemma 5.5(b), the map  $(y_1, y_2) \in Y \times Y \rightarrow \sup_{u \in U} \|f^{(l)}(y_1, u) - f^{(l)}(y_2, u)\|$  is well defined. Further using the fact that for every  $(y_1, y_2), (y'_1, y'_2) \in Y \times Y$ ,  $|\sup_{u \in U} \|f^{(l)}(y_1, u) - f^{(l)}(y_2, u)\| - \sup_{u \in U} \|f^{(l)}(y'_1, u) - f^{(l)}(y'_2, u)\|| \leq \sup_{u \in U} \|f^{(l)}(y_1, u) - f^{(l)}(y'_1, u)\| + \sup_{u \in U} \|f^{(l)}(y_2, u) - f^{(l)}(y'_2, u)\|$  and Lemma 5.5(c), we have that the map  $(y_1, y_2) \rightarrow \sup_{u \in U} \|f^{(l)}(y_1, u) - f^{(l)}(y_2, u)\|$  is continuous. Thus the map  $(y_1, y_2) \in E \rightarrow \frac{\sup_{u \in U} \|f^{(l)}(y_1, u) - f^{(l)}(y_2, u)\|}{\|y_1 - y_2\|}$  is a continuous function on a compact set  $E$  and hence achieves a maximum, say  $L_1(Y, l) \geq 0$ . Therefore for every  $(y_1, y_2) \in (Y \times Y) \cap \{(y_1, y_2) : \|y_1 - y_2\| \geq \delta(Y, l)\}$ ,  $\sup_{u \in U} \|f^{(l)}(y_1, u) - f^{(l)}(y_2, u)\| \leq L_1(Y, l)\|y_1 - y_2\|$ .

Thus from the arguments in the two preceding paragraphs we have that there exists  $L(Y, l) := \max\{L_0(Y, l), L_1(Y, l)\}$ , such that, for every  $y_1, y_2 \in Y$ ,  $\sup_{u \in U} \|f^{(l)}(y_1, u) - f^{(l)}(y_2, u)\| \leq L(Y, l)\|y_1 - y_2\|$ .

- (d) Using parts (b) and (c) of this lemma, the proof of uniqueness follows from arguments similar to that of [16, Thm. 3.4]. ■

### 5.3 Bounding procedure

In this section we show that the lower bound on the probability of the event that the iterates converge to an attracting set given that after a large number of iterations the iterates lies in a neighborhood of it depends mainly on the additive noise terms.

In order to accomplish this we first define some terms which are a measure of the distance of the linearly interpolated trajectory of recursion (5), that is  $\bar{X}$  (see (9)) to the solutions of the DI (6) over a  $T > 0$  length time interval, among others. Recall from section 3.2 that  $\mathcal{O}' \subseteq \mathbb{R}^d$ , is an open neighborhood of the attracting set  $A$  (as in assumption (A4)) with compact closure, such that  $A \subseteq \mathcal{O}' \subseteq \bar{\mathcal{O}}' \subseteq \mathcal{O}$ , where  $\mathcal{O}$  denotes the fundamental neighborhood of  $A$ . Thus we can find an  $\epsilon_0 > 0$ , such that  $N^{\epsilon_0}(\bar{\mathcal{O}}') \subseteq \mathcal{O}$  and  $N^{2\epsilon_0}(A) \subseteq \mathcal{O}'$ , where for any  $\epsilon > 0$ ,  $N^\epsilon(\cdot)$  denotes the  $\epsilon$ -neighborhood of a set. Further, since  $A$  is an attracting set for the flow of DI (6), for  $\epsilon_0 > 0$  as obtained above, there exists  $T_A > 0$ , such that for every  $x \in \mathcal{O}$ , for every  $t \geq T_A$ ,  $\Phi(t, x) \in N^{\epsilon_0}(A)$ . Throughout the rest of this paper  $\epsilon_0$  and  $T_A$  will denote the constants as obtained above.

For every  $T > 0$ , for every  $n \geq 0$ ,

**Definition 1 :** let  $\tau(n, T) := \min\{k \geq n : t(k) \geq t(n) + T\}$ , where  $t(n)$ , for every  $n \geq 0$  are as defined in section 3.1. That is  $\tau(n, T)$  denotes the first iterate such that, at least time  $T$  has elapsed since the  $n^{th}$  iteration. Further the time elapsed from iteration  $n$  to iteration  $\tau(n, T)$ , be denoted by  $\Delta(n, T)$ , that is  $\Delta(n, T) := t(\tau(n, T)) - t(n)$ . Then by the choice of our step sizes we have that  $T \leq \Delta(n, T) \leq T + 1$ .

**Definition 2 :** for every  $\omega \in \Omega$ ,  $\rho(\omega, n, T) := \inf_{\mathbf{x} \in S(T, \bar{\mathcal{O}}')} \sup_{t \in [0, T]} \|\bar{X}(\omega, t + t(n)) - \mathbf{x}(t)\|$ , where  $S(T, \bar{\mathcal{O}}')$  denotes the set of solutions of DI (6) as defined in equation (21).

**Definition 3 :** for every  $\omega \in \Omega$ , for every  $l \geq 1$ , let  $\bar{\mathbf{x}}^{(l)}(\cdot; n, T, \omega) : [0, T] \rightarrow \mathbb{R}^d$  denote the unique solution of the o.d.e.

$$\frac{dx}{dt} = f^{(l)}(x, u(t; n, T, \omega)), \quad (25)$$



with initial condition  $\bar{\mathbf{x}}^{(l)}(0; n, T, \omega) = X_n(\omega)$ , where  $u(\cdot; n, T, \omega) : [0, T] \rightarrow U$  is defined such that, for every  $t \in [0, T]$ ,  $u(t; n, T, \omega) := U_k^{(l)}(\omega)$ , where  $U_k^{(l)}$  is as in equation (19) and  $k$  is such that  $t + t(n) \in [t(k), t(k+1))$  (for a proof of existence and uniqueness of solutions to o.d.e. (25), see Lemma 5.7). It is easy to see that for every  $l \geq 1$ ,  $\bar{\mathbf{x}}^{(l)}(\cdot; n, T, \omega) \in S^{(l)}(T, X_n(\omega))$ , where  $S^{(l)}(T, X_n(\omega))$  denotes the set of solutions of DI (20), as defined in (22).

**Definition 4** : for every  $\omega \in \Omega$ , for every  $l \geq 1$ ,  $\rho_1^{(l)}(\omega, n, T) := \sup_{t \in [0, T]} \|\bar{X}(\omega, t + t(n)) - \bar{\mathbf{x}}^{(l)}(t; n, T, \omega)\|$  and  $\rho_2^{(l)}(\omega, n, T) := \inf_{\mathbf{x} \in S(T, \bar{\mathcal{O}}')} \sup_{t \in [0, T]} \|\bar{\mathbf{x}}^{(l)}(t; n, T, \omega) - \mathbf{x}(t)\|$ .

**Definition 5** : for any  $T_u \geq T_A$ , for any  $n_0 \geq 0$ , let  $\{n_m\}_{m \geq 1}$  denote a subsequence of natural numbers defined such that for every  $m \geq 0$ ,  $T_A \leq T_m := t(n_{m+1}) - t(n_m) \leq T_u$ .

Now we collect sample paths of interest using the quantities  $\rho$ ,  $\rho_1^{(l)}$  and  $\rho_2^{(l)}$ . The next lemma summarizes results in this regard.

**Lemma 5.8** *For every  $T_u \geq T_A$ , for every  $n_0 \geq 0$ , for every  $l \geq 1$ , for every event  $E \in \mathcal{F}_{n_0}$ , such that  $E \subseteq \{\omega : X_{n_0}(\omega) \in \mathcal{O}'\}$ , for every  $\{n_m\}_{m \geq 1}$  as in **Definition 5**,*

(a) *for every  $M \geq 0$ ,*

$$E \cap \left( \bigcap_{m=0}^M \left\{ \omega \in \Omega : \rho_1^{(l)}(\omega, n_m, T_m) + \rho_2^{(l)}(\omega, n_m, T_m) < \epsilon_0 \right\} \right) \subseteq E \cap \left( \bigcap_{m=0}^M \left\{ \omega \in \Omega : \rho(\omega, n_m, T_m) < \epsilon_0 \right\} \right) \\ \subseteq \left\{ \omega \in \Omega : X_{n_{M+1}}(\omega) \in \mathcal{O}' \right\},$$

(b)

$$\mathbb{P} \left( E \cap \left( \bigcap_{m \geq 0} \left\{ \omega \in \Omega : \rho_1^{(l)}(\omega, n_m, T_m) + \rho_2^{(l)}(\omega, n_m, T_m) < \epsilon_0 \right\} \right) \right) \leq \mathbb{P}(E \cap (\bigcap_{m \geq 0} \{ \omega \in \Omega : \rho(\omega, n_m, T_m) < \epsilon_0 \})) \\ \leq \mathbb{P}(E \cap \{ \omega \in \Omega : X_n(\omega) \rightarrow A \text{ as } n \rightarrow \infty \}),$$

where,  $\{T_m\}_{m \geq 0}$  is as in **Definition 5**.

**Proof** Fix  $n_0 \geq 0$ ,  $l \geq 1$  and  $E \in \mathcal{F}_{n_0}$ , such that  $E \subseteq \{\omega \in \Omega : X_{n_0}(\omega) \in \mathcal{O}'\}$ .

(a) For every  $m \geq 0$ , for every  $\omega \in \Omega$ , from **Definition 2** and **Definition 4**, it is clear that,

$$\rho(\omega, n_m, T_m) \leq \rho_1^{(l)}(\omega, n_m, T_m) + \rho_2^{(l)}(\omega, n_m, T_m),$$

from which we get that for every  $m \geq 0$ ,

$$\{\omega \in \Omega : \rho_1^{(l)}(\omega, n_m, T_m) + \rho_2^{(l)}(\omega, n_m, T_m) < \epsilon_0\} \subseteq \{\omega \in \Omega : \rho(\omega, n_m, T_m) < \epsilon_0\}.$$

Therefore,

$$E \cap \left( \bigcap_{m=0}^M \left\{ \omega \in \Omega : \rho_1^{(l)}(\omega, n_m, T_m) + \rho_2^{(l)}(\omega, n_m, T_m) < \epsilon_0 \right\} \right) \subseteq E \cap \left( \bigcap_{m=0}^M \left\{ \omega \in \Omega : \rho(\omega, n_m, T_m) < \epsilon_0 \right\} \right).$$

The proof of the second inclusion follows from induction. Fix  $M = 0$  and  $\omega \in E \cap \{\omega \in \Omega : \rho(\omega, n_0, T_0) < \epsilon_0\}$ . Then  $X_{n_0}(\omega) \in \mathcal{O}'$ . Since  $T_0 \geq T_A$ , we have that for every  $\mathbf{x} \in S(T_0, \bar{\mathcal{O}}')$ ,  $\mathbf{x}(T_0) \in N^{\epsilon_0}(A)$ . Further, since  $\rho(\omega, n, T_0) < \epsilon_0$  and by Lemma 5.6(d), we get that there exists  $\mathbf{x} \in S(T_0, \bar{\mathcal{O}}')$ , such that  $\|\bar{X}(\omega, t(n_1)) - \mathbf{x}(T_0)\| = \|X_{n_1}(\omega) - \mathbf{x}(T_0)\| < \epsilon_0$  and hence  $X_{n_1}(\omega) \in N^{2\epsilon_0}(A) \subseteq \mathcal{O}'$ . Therefore  $\omega \in \{\omega \in \Omega : X_{n_1}(\omega) \in \mathcal{O}'\}$ . Thus the inclusion is true for  $M = 0$ . Suppose the inclusion is true for some  $M > 0$ . Let  $\omega \in E \cap \left( \bigcap_{m=0}^{M+1} \left\{ \omega \in \Omega : \rho(\omega, n_m, T_m) < \epsilon_0 \right\} \right)$ . Since the inclusion is true for  $M$ , we have that  $X_{n_{M+1}}(\omega) \in \mathcal{O}'$ . Now by arguments exactly same as those for the base case (that is for  $M = 0$ ) we get that  $X_{n_{M+2}}(\omega) \in \mathcal{O}'$ . Therefore the inclusion is true for  $M + 1$ .

(b) The first inequality follows from part (a) of this lemma. We shall provide a proof of the second inequality. Let  $\omega \in E \cap (\bigcap_{m \geq 0} \{\omega \in \Omega : \rho(\omega, n_m, T_m) < \epsilon_0\})$ . Then by part (a) of this lemma we have that for every  $m \geq 0$ ,  $X_{n_m}(\omega) \in \mathcal{O}'$ . Since  $\bar{\mathcal{O}}'$  is compact, by Lemma 5.6(d), we have that,  $S(T_u, \bar{\mathcal{O}}')$  is a compact subset of  $\mathcal{C}([0, T_u], \mathbb{R}^d)$ , and hence there exists  $C(\bar{\mathcal{O}}', T_m) > 0$  such that,  $\sup_{\mathbf{x} \in S(T_u, \bar{\mathcal{O}}')} \sup_{t \in [0, T_u]} \|\mathbf{x}(t)\| \leq C(\bar{\mathcal{O}}', T_m)$ . Further since for every  $m \geq 0$ ,  $T_m \leq T_u$ , we get that  $\sup_{\mathbf{x} \in S(T_m, \bar{\mathcal{O}}')} \sup_{t \in [0, T_m]} \|\mathbf{x}(t)\| \leq \sup_{\mathbf{x} \in S(T_u, \bar{\mathcal{O}}')} \sup_{t \in [0, T_u]} \|\mathbf{x}(t)\| \leq C(\bar{\mathcal{O}}', T_m)$ . By our choice of  $\omega$ , we have that for every  $m \geq 0$ ,  $\rho(\omega, n_m, T_m) < \epsilon_0$  and by **Definition 2**, we

get that for every  $m \geq 0$ ,  $\sup_{t \in [0, T_m]} \|\bar{X}(\omega, t + t(n_m))\| \leq C(\bar{\mathcal{O}}', T_m) + \epsilon_0$ . Therefore  $\omega$ , is such that  $\sup_{n \geq 0} \|X_n(\omega)\| < \infty$  and for every  $m \geq 0$ ,  $X_{n_m}(\omega) \in \mathcal{O}'$ . Thus  $\lambda(\omega)$  (see equation (10) for definition), is non-empty, compact and  $\lambda(\omega) \cap \bar{\mathcal{O}}' \subseteq \lambda(\omega) \cap B(A) \neq \emptyset$ , where  $B(A)$  denotes the basin of attraction of the attracting set  $A$ . By Theorem 3.1(a), we have that for almost every  $\omega$  in  $E \cap (\cap_{m \geq 0} \{\omega \in \Omega : \rho(\omega, n_m, T_m) < \epsilon_0\})$  the iterates converge to the attracting set  $A$ . Therefore we get that  $\mathbb{P}(E \cap (\cap_{m \geq 0} \{\omega \in \Omega : \rho(\omega, n_m, T_m) < \epsilon_0\})) \leq \mathbb{P}(E \cap \{\omega \in \Omega : X_n(\omega) \rightarrow A \text{ as } n \rightarrow \infty\})$ . ■

The quantity  $\rho_1^{(l)}$  as in **Definition 4**, captures the difference between the linearly interpolated trajectory of recursion (5) and the solution of the o.d.e. (25) over a  $T > 0$  length time interval. This difference can be shown to comprise of two components namely, the error due to discretization and the error due to additive noise terms. By the step size assumption, that is (A2), we know that the step sizes are converging to zero. Hence intuition suggests that after a large number of iterations have elapsed the discretization error must be negligible and the contribution to the difference term  $\rho_1^{(l)}$  is mainly due to the additive noise terms. The following is made precise in the lemma below. A brief outline of the proof of this lemma which follows from Lemma 5.7(c) and [3, Ch. 2, Lemma 1], is presented in Appendix B.

**Lemma 5.9** *For every  $l \geq 1$ , for every  $T_u \geq T_A$ , there exists  $N'_0 \geq 1$ , such that for every  $n_0 \geq N'_0$ , for every  $E \in \mathcal{F}_{n_0}$  such that,  $E \subseteq \{\omega \in \Omega : X_{n_0}(\omega) \in \mathcal{O}'\}$ , for every sequence  $\{n_m\}_{m \geq 0}$  as in **Definition 5**, for every  $m \geq 0$ , we have,*

$$\mathbb{P}\left(\mathcal{B}_{m-1}^{(l)} \cap \{\omega \in \Omega : \rho_1^{(l)}(\omega, n_m, T_m) \geq \frac{\epsilon_0}{2}\}\right) \leq \mathbb{P}\left(\{\omega \in \Omega : \max_{n_m \leq j \leq n_{m+1}} \|\zeta_j(\omega) - \zeta_{n_m}(\omega)\| \geq \frac{\epsilon_0}{4K_0(T_u)}\} \cap \mathcal{B}_{m-1}^{(l)}\right),$$

where,

- $\mathcal{B}_{-1}^{(l_0)} := E$  and for every  $M \geq 0$ ,  $\mathcal{B}_M^{(l_0)} := E \cap \left(\cap_{m=0}^M \{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_m, T_m) + \rho_2^{(l_0)}(\omega, n_m, T_m) < \epsilon_0\}\right)$ ,
- for every  $j \geq 1$ ,  $\zeta_j := \sum_{n=0}^{j-1} a(n)M_{n+1}$ , where  $\{M_n\}_{n \geq 1}$  denote the additive noise terms as defined in assumption (A3),
- $\{T_m\}_{m \geq 0}$  is as in **Definition 5** and  $K_0(T_u) > 0$  is a positive constant increasing in  $T_u$

Suppose event  $E$  as in the lemma above occurs with some positive probability. Then the next lemma says that the lower bound of  $\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \rightarrow A \text{ as } n \rightarrow \infty\} | E)$  depends mainly on the additive noise terms for  $n_0$  large.

**Lemma 5.10** *For every  $T_u \geq T_A$ , there exists  $l_0 \geq 1$  and  $N'_0 \geq 1$ , such that for every  $n_0 \geq N'_0$ , for every  $E \in \mathcal{F}_{n_0}$  such that,  $E \subseteq \{\omega \in \Omega : X_{n_0}(\omega) \in \mathcal{O}'\}$  and  $\mathbb{P}(E) > 0$ , for every sequence  $\{n_m\}_{m \geq 0}$  as in **Definition 5**, we have,*

$$\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \rightarrow A \text{ as } n \rightarrow \infty\} | E) \geq 1 - \sum_{m=0}^{\infty} \mathbb{P}\left(\max_{n_m \leq j \leq n_{m+1}} \|\zeta_j - \zeta_{n_m}\| \geq \frac{\epsilon_0}{4K_0(T_u)} | \mathcal{B}_{m-1}^{(l_0)}\right), \quad (26)$$

where, the sequence of events  $\{\mathcal{B}_m^{(l_0)}\}_{m \geq -1}$ , the sequence of random vectors  $\{\zeta_j\}_{j \geq 1}$  and the constant  $K_0(T_u)$  are as defined in Lemma 5.9.

**Proof** By Lemma 5.6(f), we get that there exists  $l_0 \geq 1$  (depending on  $\bar{\mathcal{O}}'$ ,  $T_u$  and  $\epsilon_0$ ) such that for every  $\mathbf{x}^{(l_0)} \in S^{(l_0)}(T_u, \bar{\mathcal{O}}')$ , there exists  $\mathbf{x} \in S(T_u, \bar{\mathcal{O}}')$  such that  $\sup_{t \in [0, T_u]} \|\mathbf{x}^{(l_0)}(t) - \mathbf{x}(t)\| < \frac{\epsilon_0}{2}$ . Further by Lemma 5.8(a) and definition of  $E$ , we get that for every  $m \geq 0$ ,  $\mathcal{B}_{m-1}^{(l_0)} \subseteq \{\omega \in \Omega : X_{n_m}(\omega) \in \mathcal{O}'\}$ . Therefore, for every  $\omega \in \mathcal{B}_{m-1}^{(l_0)}$ ,  $\bar{\mathbf{x}}^{(l_0)}(\cdot; n_m, T_m, \omega) \in S^{(l_0)}(T_m, \bar{\mathcal{O}}')$  and

$$\begin{aligned} \rho_2^{(l_0)}(\omega, n_m, T_m) &= \inf_{\mathbf{x} \in S(T_m, \bar{\mathcal{O}}')} \sup_{t \in [0, T_m]} \|\bar{\mathbf{x}}^{(l_0)}(t; n_m, T_m, \omega) - \mathbf{x}(t)\| \\ &\leq \inf_{\mathbf{x} \in S(T_u, \bar{\mathcal{O}}')} \sup_{t \in [0, T_u]} \|\bar{\mathbf{x}}^{(l_0)}(t; n, T_u, \omega) - \mathbf{x}(t)\| \end{aligned} \quad (27)$$

$$< \frac{\epsilon_0}{2}, \quad (28)$$

where (27) follows from the fact that  $T_m \leq T_u$  and (28) follows from our choice of  $l_0$  and **Definition 3**. Therefore for every  $m \geq -1$ ,

$$\mathcal{B}_m^{(l_0)} \cap \{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_{m+1}, T_m) + \rho_2^{(l_0)}(\omega, n_{m+1}, T_m) \geq \epsilon_0\} \subseteq \mathcal{B}_m^{(l_0)} \cap \{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_{m+1}, T_m) \geq \frac{\epsilon_0}{2}\},$$

and hence,

$$\mathbb{P}\left(\{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_{m+1}, T_m) + \rho_2^{(l_0)}(\omega, n_{m+1}, T_m) \geq \epsilon_0\} | \mathcal{B}_m^{(l_0)}\right) \leq \mathbb{P}\left(\{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_{m+1}, T_m) \geq \frac{\epsilon_0}{2}\} | \mathcal{B}_m^{(l_0)}\right). \quad (29)$$

By Lemma 5.9, we know that there exists  $N'_0 \geq 1$  such that, for every  $n_0 \geq N'_0$ , for every  $m \geq 0$ ,

$$\mathbb{P}\left(\mathcal{B}_{m-1}^{(l_0)} \cap \{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_m, T_m) \geq \frac{\epsilon_0}{2}\}\right) \leq \mathbb{P}\left(\mathcal{B}_{m-1}^{(l_0)} \cap \{\omega \in \Omega : \max_{n_m \leq j \leq n_{m+1}} \|\zeta_j(\omega) - \zeta_{n_m}(\omega)\| \geq \frac{\epsilon_0}{4K_0(T_u)}\}\right),$$

from which it follows that,

$$\mathbb{P}\left(\{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_m, T_m) \geq \frac{\epsilon_0}{2}\} | \mathcal{B}_{m-1}^{(l_0)}\right) \leq \mathbb{P}\left(\{\omega \in \Omega : \max_{n_m \leq j \leq n_{m+1}} \|\zeta_j(\omega) - \zeta_{n_m}(\omega)\| \geq \frac{\epsilon_0}{4K_0(T_u)}\} | \mathcal{B}_{m-1}^{(l_0)}\right). \quad (30)$$

For  $l_0 \geq 1$  as obtained above and for  $n_0 \geq N'_0$ , we have that,

$$\mathbb{P}(X_n \rightarrow A \text{ as } n \rightarrow \infty | E) \geq \mathbb{P}\left(\cap_{m \geq 0} \{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_m, T_m) + \rho_2^{(l_0)}(\omega, n_m, T_m) < \epsilon_0\} | E\right) \quad (31)$$

$$\begin{aligned} &= 1 - \mathbb{P}\left(\cup_{m \geq 0} \{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_m, T_m) + \rho_2^{(l_0)}(\omega, n_m, T_m) \geq \epsilon_0\} | E\right) \\ &= 1 - \mathbb{P}\left(\{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_0, T_m) + \rho_2^{(l_0)}(\omega, n_0, T_m) \geq \epsilon_0\} | \mathcal{B}_{-1}^{(l_0)}\right) \\ &\quad - \sum_{m=1}^{\infty} \mathbb{P}\left(\{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_m, T_m) + \rho_2^{(l_0)}(\omega, n_m, T_m) \geq \epsilon_0\} | \mathcal{B}_{m-1}^{(l_0)}\right) \mathbb{P}\left(\mathcal{B}_{m-1}^{(l_0)} | \mathcal{B}_{-1}^{(l_0)}\right) \end{aligned} \quad (32)$$

$$\geq 1 - \sum_{m=0}^{\infty} \mathbb{P}\left(\{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_m, T_m) + \rho_2^{(l_0)}(\omega, n_m, T_m) \geq \epsilon_0\} | \mathcal{B}_m^{(l_0)}\right), \quad (33)$$

where, (31) follows from Lemma 5.8(b), (32) follows from the observation that,

$$\begin{aligned} &\left(\cup_{m \geq 0} \{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_m, T_m) + \rho_2^{(l_0)}(\omega, n_m, T_m) \geq \epsilon_0\}\right) \cap E = \\ &\quad \cup_{m \geq 0} \left(\{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_m, T_m) + \rho_2^{(l_0)}(\omega, n_m, T_m) \geq \epsilon_0\} \cap \mathcal{B}_m^{(l_0)}\right), \end{aligned}$$

(where the union in R.H.S. is disjoint) and (33) follows from the fact that for every  $m \geq 0$ ,  $\mathbb{P}(\mathcal{B}_m^{(l_0)} | \mathcal{B}_{-1}^{(l_0)}) \leq 1$ . Using (29) and (30) in (33), we get that there exists  $l_0 \geq 1$  and  $N'_0 \geq 1$ , such that for every  $n_0 \geq N'_0$ , for every  $E \in \mathcal{F}_{n_0}$  such that  $E \subseteq \{\omega \in \Omega : X_{n_0}(\omega) \in \mathcal{O}'\}$  and  $\mathbb{P}(E) > 0$ ,

$$\begin{aligned} \mathbb{P}(X_n \rightarrow A \text{ as } n \rightarrow \infty | E) &\geq 1 - \sum_{m=0}^{\infty} \mathbb{P}\left(\{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_m, T_m) + \rho_2^{(l_0)}(\omega, n_m, T_m) \geq \epsilon_0\} | \mathcal{B}_{m-1}^{(l_0)}\right) \\ &\geq 1 - \sum_{m=0}^{\infty} \mathbb{P}\left(\{\omega \in \Omega : \rho_1^{(l_0)}(\omega, n_m, T_m) \geq \frac{\epsilon_0}{2}\} | \mathcal{B}_{m-1}^{(l_0)}\right) \\ &\geq 1 - \sum_{m=0}^{\infty} \mathbb{P}\left(\{\omega \in \Omega : \max_{n_m \leq j \leq n_{m+1}} \|\zeta_j(\omega) - \zeta_{n_m}(\omega)\| \geq \frac{\epsilon_0}{4K_0(T_u)}\} | \mathcal{B}_{m-1}^{(l_0)}\right). \quad \blacksquare \end{aligned}$$

## 5.4 Review of the probability bounding procedure for the additive noise terms

In this section we fix  $l_0$  and  $n_0 \geq N_0$ , where  $l_0$  and  $N_0$  are as in Lemma 5.10 and provide an upper bound for  $\mathbb{P}\left(\{\omega \in \Omega : \max_{n_m \leq j \leq n_{m+1}} \|\zeta_j(\omega) - \zeta_{n_m}(\omega)\| \geq \frac{\epsilon_0}{4K_0(T_u)}\} | \mathcal{B}_{m-1}^{(l_0)}\right)$ , for every  $m \geq 0$ . The proof of the bounding procedure is similar to that of [3, Ch. 4, Lemma. 10] and we provide a brief outline here for the sake of completeness.

- (a) From recursion (5), we have that for every  $m \geq 0$ , for every  $n_m \leq j \leq n_{m+1} - 1$ , for every  $\omega \in \Omega$ , there exists  $V_j(\omega) \in F(X_j(\omega))$ , such that,

$$X_{j+1}(\omega) - X_j(\omega) - a(n)M_{j+1}(\omega) = a(n)V_j(\omega).$$

By assumption (A1)(ii), we know that  $\|V_j(\omega)\| \leq K(1 + \|X_j(\omega)\|)$  and hence for  $n_m \leq j \leq n_{m+1} - 1$ ,

$$\|X_{j+1}(\omega)\| \leq \|X_j(\omega)\|(1 + a(j)K) + a(j)K + a(j)\|M_{j+1}\|.$$

Further by assumption (A3), we get that, for every  $m \geq 0$ , for almost every  $\omega \in \Omega$ , for  $n_m \leq j \leq n_{m+1} - 1$ ,

$$\|X_{j+1}(\omega)\| \leq \|X_j(\omega)\|(1 + 2a(j)K) + 2a(j)K.$$

Now by arguments as in [3, Lemma 9], we get that, for every  $m \geq 0$ , for almost every  $\omega \in \Omega$ , for  $n_m \leq j \leq n_{m+1}$ ,

$$\|X_j(\omega)\| \leq e^{2KT_u}(\|X_{n_m}(\omega)\| + 2KT_u). \quad (34)$$

- (b) Clearly  $\{\zeta_j - \zeta_{n_m}, \mathcal{F}_j\}_{n_m \leq j \leq n_{m+1}}$  is a martingale. By (34) and (A3), we get that for  $n_m \leq j < n_{m+1}$ ,  $\|\zeta_{j+1} - \zeta_j\| = \|a(j)M_{j+1}\| \leq a(j)K(1 + \|X_j\|) \leq a(j)K(1 + e^{2KT_u}(1 + 2KT_u\|X_{n_m}\|))$ . Since for every  $\omega \in \mathcal{B}_{m-1}^{(l_0)}$ ,  $X_{n_m}(\omega) \in \mathcal{O}'$  (whose closure is compact), there exists a  $C > 0$ , such that  $\|X_{n_m}(\omega)\| \leq C$ . Therefore for every  $m \geq 0$ , for every  $\omega \in \mathcal{B}_{m-1}^{(l_0)}$ , for every  $n_m \leq j < n_{m+1}$ ,  $\|\zeta_{j+1} - \zeta_j\| \leq a(j)K(1 + e^{2KT_u}(1 + 2KT_uC))$ . Thus applying the concentration inequality for martingales, by arguments exactly the same as in the proof of [3, Lemma 10], we get that for every  $m \geq 0$ ,

$$\mathbb{P}\left(\{\omega \in \Omega : \max_{n_m \leq j \leq n_{m+1}} \|\zeta_j(\omega) - \zeta_{n_m}(\omega)\| \geq \frac{\epsilon_0}{4K_0(T_u)}\} | \mathcal{B}_{m-1}^{(l_0)}\right) \leq 2de^{-\tilde{K}/(b(n_m) - b(n_{m+1}))} \quad (35)$$

where,  $\tilde{K} := \epsilon_0^2 / (32(K_0(T_u))^2 dK(1 + e^{2KT_u}(1 + 2KT_uC)))$ .

**Proof of Theorem 3.2:** Let  $l_0 \geq 1$  and  $N'_0$  be as in Lemma 5.10. By definition of  $b(\cdot)$ , we get that there exists  $N''_0 \geq 1$ , such that for every  $n \geq N''_0$ ,  $b(n) < \tilde{K}$ . Define  $N_0 := \max\{N'_0, N''_0\}$ . Let  $n_0 \geq N_0$  and  $\{n_m := \tau(n_{m-1}, T_A)\}_{m \geq 1}$ .  $\{n_m\}_{m \geq 1}$  as defined satisfies the conditions mentioned in **Definition 5**. Then by Lemma 5.10 and (35), we get that for  $n_0 \geq N_0$ ,

$$\mathbb{P}(X_n \rightarrow A \text{ as } n \rightarrow \infty | E) \geq 1 - 2d \sum_{m=0}^{\infty} e^{-\tilde{K}/(b(n_m) - b(n_{m+1}))}. \quad (36)$$

We know that  $e^{-\tilde{K}/x}/x \rightarrow 0$  as  $x \rightarrow 0$  and increases with  $x$  for  $0 < x < \tilde{K}$ . Therefore by our choice of  $n_0$ , we get that,

$$\frac{e^{-\tilde{K}/(b(n_m) - b(n_{m+1}))}}{b(n_m) - b(n_{m+1})} \leq \frac{e^{-\tilde{K}/b(n_0)}}{b(n_0)},$$

from which it follows that for every  $m \geq 0$ ,  $e^{-\frac{\tilde{K}}{b(n_m) - b(n_{m+1})}} \leq (b(n_m) - b(n_{m+1})) \frac{e^{-\frac{\tilde{K}}{b(n_0)}}}{b(n_0)}$ . Substituting the above in (36), we get that for every  $n_0 \geq N_0$ ,

$$\begin{aligned} \mathbb{P}(X_n \rightarrow A \text{ as } n \rightarrow \infty | E) &\geq 1 - 2d \sum_{m=0}^{\infty} e^{-\frac{\tilde{K}}{b(n_m) - b(n_{m+1})}} \\ &\geq 1 - 2d \sum_{m=0}^{\infty} (b(n_m) - b(n_{m+1})) \frac{e^{-\tilde{K}/b(n_0)}}{b(n_0)} \\ &= 1 - 2d \frac{e^{-\tilde{K}/b(n_0)}}{b(n_0)} \sum_{m=0}^{\infty} (b(n_m) - b(n_{m+1})) \\ &= 1 - 2de^{-\tilde{K}/b(n_0)}. \quad \blacksquare \end{aligned}$$

## 6 Proof of finite resets theorem (Thm. 4.1)

From the definition of  $\chi_n$  in equation (12), we know that the  $\chi_n$  takes the value one if there is a reset of the  $n^{\text{th}}$  iterate and is zero otherwise. Therefore  $\sum_{n=1}^{\infty} \chi_n$  denotes the total number of resets.

Suppose the event  $\{\sum_{n=1}^{\infty} \chi_n \geq k\}$  has zero probability for some  $k \geq 1$ . Then for  $k \geq 1$ , such that  $\mathbb{P}(\sum_{n=1}^{\infty} \chi_n \geq k) = 0$ , we have  $\mathbb{P}(\sum_{n=1}^{\infty} \chi_n < k) = 1$ , from which Theorem 4.1 trivially follows. Therefore without loss of generality assume  $\mathbb{P}(\sum_{n=1}^{\infty} \chi_n \geq k) > 0$ , for every  $k \geq 1$ .

For every  $k \geq 0$ , let  $G_k$  denote the event that there are at most  $k$  resets and  $G_\infty$  denote the event that there are finitely many resets. That is, for every  $k \geq 0$ ,  $G_k := \{\sum_{n=1}^\infty \chi_n \leq k\}$  and  $G_\infty := \{\sum_{n=1}^\infty \chi_n < \infty\}$ . Then it is clear that, for every  $k \geq 1$ ,  $G_k \subseteq G_{k+1}$  and  $G_\infty = \cup_{k \geq 0} G_k$ . Therefore  $\lim_{k \rightarrow \infty} \mathbb{P}(G_k)$  exists and  $\mathbb{P}(G_\infty) = \lim_{k \rightarrow \infty} \mathbb{P}(G_k)$ . For any  $k \geq 1$ ,

$$\mathbb{P}(G_k) = \mathbb{P}\left(\sum_{n=1}^\infty \chi_n \leq k\right) = \mathbb{P}\left(\left\{\sum_{n=1}^\infty \chi_n \leq k-1\right\} \cup \left\{\sum_{n=1}^\infty \chi_n = k\right\}\right) = \mathbb{P}(G_{k-1}) + \mathbb{P}\left(\sum_{n=1}^\infty \chi_n = k\right). \quad (37)$$

The event  $\{\sum_{n=1}^\infty \chi_n = k\}$  can be written as a disjoint union of events as below. For every  $k \geq 1$

$$\left\{\sum_{n=1}^\infty \chi_n = k\right\} = \cup_{n_0 \geq 1} \left[ \left\{\sum_{n=1}^{n_0-1} \chi_n = k-1\right\} \cap \{\chi_{n_0} = 1\} \cap \left\{\sum_{n=n_0+1}^\infty \chi_n = 0\right\} \right], \quad (38)$$

where,  $\{\sum_{n=1}^0 \chi_n = k-1\} := \Omega$ . Let  $J(k) := \{n_0 \geq 1 : \mathbb{P}\left(\left\{\sum_{n=1}^{n_0-1} \chi_n = k-1\right\} \cap \{\chi_{n_0} = 1\}\right) > 0\}$ . Then for every  $k \geq 1$ ,

- (a) By arguments in the second paragraph of this section we have that  $\mathbb{P}(G_{k-1}^c) = \mathbb{P}(\sum_{n=1}^\infty \chi_n \geq k) > 0$ . Further the event  $\{\sum_{n=1}^\infty \chi_n \geq k\}$  can be written as a disjoint union of events as below.

$$\left\{\sum_{n=1}^\infty \chi_n \geq k\right\} = \cup_{n_0 \geq 1} \left[ \left\{\sum_{n=1}^{n_0-1} \chi_n = k-1\right\} \cap \{\chi_{n_0} = 1\} \right], \quad (39)$$

from which it follows that,

$$0 < \mathbb{P}\left(\left\{\sum_{n=1}^\infty \chi_n \geq k\right\}\right) = \sum_{n_0=1}^\infty \mathbb{P}\left(\left\{\sum_{n=1}^{n_0-1} \chi_n = k-1\right\} \cap \{\chi_{n_0} = 1\}\right). \quad (40)$$

Therefore  $J(k) \neq \emptyset$ .

- (b)  $\min\{n_0 \in J(k)\} \geq k$ , since there cannot be  $k$  resets in less than  $k$  iterations.

From (38) and definition of  $J(k)$ , we have that for every  $k \geq 1$ ,

$$\mathbb{P}\left(\sum_{n=1}^\infty \chi_n = k\right) = \sum_{n_0 \in J(k)} \mathbb{P}\left(\sum_{n=n_0+1}^\infty \chi_n = 0 \mid \sum_{n=1}^{n_0-1} \chi_n = k-1, \chi_{n_0} = 1\right) \mathbb{P}\left(\sum_{n=1}^{n_0-1} \chi_n = k-1, \chi_{n_0} = 1\right). \quad (41)$$

**Step 1 (Obtaining  $\mathcal{O}'$ ,  $\epsilon_0$  and  $T_A$ ) :** By (A4)', we have that  $A$  is a globally attracting set of DI (6). Let  $\tilde{r} > 0$  be such that  $A \subseteq \tilde{r}\tilde{U}$ . By definition of a globally attracting set and [2, Lemma 3.13], we get that for any  $r \geq \tilde{r}$ ,  $r\tilde{U}$  is a fundamental neighborhood of  $A$ . Let  $k_1 \geq 1$  be such that  $r_{k_1} \geq \tilde{r}$ . Set the fundamental neighborhood  $\mathcal{O} := r_{k_1+1}\tilde{U}$  and  $\mathcal{O}' := r_{k_1}\tilde{U}$ . Obtain  $\epsilon_0 > 0$  and  $T_A > 0$  as in section 5.3. That is  $\epsilon_0 > 0$  is such that  $N^{2\epsilon_0}(A) \subseteq \mathcal{O}' \subseteq N^{\epsilon_0}(\tilde{\mathcal{O}}') \subseteq \mathcal{O}$  and  $T_A > 0$ , is such that for every  $x \in \mathcal{O}$ , for every  $t \geq T_A$ ,  $\Phi(t, x) \in N^{\epsilon_0}(A)$ .

**Step 2 (Obtaining  $\{n_m\}_{m \geq 1}$  as in Definition 5) :** Clearly there exists  $k_2 \geq 1$ , such that for every  $k \geq k_2$ ,  $T_A \leq 2^k T_W$ . For any  $n_0 \geq 1$ , for every  $m \geq 1$ , define  $n_m := n_{2^{k_2}, m-1}$ , where for every  $1 \leq j \leq 2^{k_2}$ ,  $n_{j, m-1} := \tau(n_{j-1, m-1}, T_W)$  with  $n_{0, m-1} := n_{m-1}$  and  $\tau(\cdot, \cdot)$  is as defined in Definition 1. Therefore for every  $m \geq 1$ ,  $T_{m-1} := t(n_m) - t(n_{m-1}) = \sum_{j=0}^{2^{k_2}-1} \Delta(n_{j, m-1}, T_W)$ , where  $\Delta(\cdot, \cdot)$  is as defined in Definition 1. Thus for every  $m \geq 0$ ,  $T_A \leq 2^{k_2} T_W \leq T_m \leq 2^{k_2} T_W + 2^{k_2}$  and hence  $T_u = 2^{k_2} T_W + 2^{k_2}$ .

**Step 3 (Redefining trajectories) :** Define  $\tilde{X}$ , as defined in (9), with the iterates  $\{X_n\}_{n \geq 0}$  (iterates before reset check) generated by Algorithm 1. For every  $n \geq 1$ , define  $\tilde{X}(\cdot, \cdot; n) : \Omega \times [t(n), \infty) \rightarrow \mathbb{R}^d$  such that for every  $(\omega, t) \in \Omega \times [t(n), t(n+1))$ ,

$$\tilde{X}(\omega, t; n) := \left(\frac{t - t(n)}{t(n+1) - t(n)}\right) X_{n+1}(\omega) + \left(\frac{t(n+1) - t}{t(n+1) - t(n)}\right) X'_n(\omega), \quad (42)$$

and for every  $(\omega, t) \in [t(n+1), \infty)$ ,  $\tilde{X}(\omega, t; n) = \bar{X}(\omega, t)$ .

**Step 4 (Obtaining parameters) :** By arguments exactly same as the ones used to obtain (19), we get that, for every  $l \geq 1$ , for every  $n \geq 0$ , there exists a  $U$ -valued random variable on  $\Omega$ , say  $\tilde{U}_n^{(l)}$  such that, for every  $\omega \in \Omega$ ,

$$X_{n+1}(\omega) - X'_n(\omega) - a(n)M_{n+1}(\omega) = a(n)f^{(l)}(X'_n(\omega), \tilde{U}_n^{(l)}(\omega)). \quad (43)$$

**Step 5 (Redefining distance measures) :** For every  $\omega \in \Omega$ , for every  $n \geq n' \geq 1$ , for every  $T > 0$ , for every  $l \geq 1$ ,

(a) let  $\tilde{\mathbf{x}}^{(l)}(\cdot; n, n', T, \omega) : [0, T] \rightarrow \mathbb{R}^d$  denote the unique solution of the o.d.e.

$$\frac{dx}{dt} = f^{(l)}(x, \tilde{u}(t; n, n', T, \omega)), \quad (44)$$

with initial condition  $\tilde{\mathbf{x}}^{(l)}(0; n, n', T, \omega) = X'_{n'}(\omega)$ , where  $\tilde{u}(\cdot; n, n', T, \omega) : [0, T] \rightarrow U$  is defined such that, for every  $t \in [0, T]$ ,  $\tilde{u}(t; n, n', T, \omega) := \tilde{U}_k^{(l)}(\omega)$ , where  $\tilde{U}_k^{(l)}$  is as in equation (43) and  $k$  is such that  $t + t(n) \in [t(k), t(k+1))$  (for a proof of existence and uniqueness of solutions to o.d.e. (44), see Lemma 5.7). It is easy to see that for every  $l \geq 1$ ,  $\tilde{\mathbf{x}}^{(l)}(\cdot; n, n', T, \omega) \in S^{(l)}(T, X'_{n'}(\omega))$ , the set of solutions of DI (20), as defined in (22).

(b) define,

- (1)  $\tilde{\rho}(\omega, n, n', T) := \inf_{\mathbf{x} \in S(T, \bar{\mathcal{O}}')} \sup_{t \in [0, T]} \|\tilde{X}(\omega, t + t(n); n') - \mathbf{x}(t)\|$ ,
- (2)  $\tilde{\rho}_1^{(l)}(\omega, n, n', T) := \sup_{t \in [0, T]} \|\tilde{X}(\omega, t + t(n); n') - \tilde{\mathbf{x}}^{(l)}(t; n, n', T, \omega)\|$ ,
- (3)  $\tilde{\rho}_2^{(l)}(\omega, n, n', T) := \inf_{\mathbf{x} \in S(T, \bar{\mathcal{O}}')} \sup_{t \in [0, T]} \|\tilde{\mathbf{x}}^{(l)}(t; n, n', T, \omega) - \mathbf{x}(t)\|$ .

**Step 6 (Collecting sample paths) :** Fix  $k > \max\{k_1, k_2\}$  and  $n_0 \in J(k)$ . By our definition of  $\mathcal{F}_{n_0}$  (see section 4), we have that  $E(k, n_0) := \{\sum_{n=1}^{n_0-1} \chi_n = k-1, \chi_{n_0} = 1\} \in \mathcal{F}_{n_0}$  and is contained in  $\{X'_{n_0}(\omega) \in \mathcal{O}'\}$ . Given that there has been a reset at index  $n_0$ , the next reset check is performed by Algorithm 1 at the iteration index  $n_{2^k-k_2}$ . So for  $n_0 + 1 \leq j < n_{2^k-k_2}$ ,  $X_j(\omega) = X'_j(\omega)$ . From arguments exactly the same as Lemma 5.8(a), we get that,

$$\begin{aligned} E(k, n_0) \cap \left( \cap_{m=0}^{2^{k-k_2}-1} \{\tilde{\rho}_1^{(l)}(\omega, n_m, n_0, T_m) + \tilde{\rho}_2^{(l)}(\omega, n_m, n_0, T_m) < \epsilon_0\} \right) \\ \subseteq E(k, n_0) \cap \left( \cap_{m=0}^{2^{k-k_2}-1} \{\tilde{\rho}(\omega, n_m, n_0, T_m) < \epsilon_0\} \right) \\ \subseteq \{\omega \in \Omega : X_{n_{2^k-k_2}}(\omega) \in \mathcal{O}'\} \\ \subseteq \{\omega \in \Omega : X_{n_{2^k-k_2}}(\omega) = X'_{n_{2^k-k_2}}(\omega)\} \\ \subseteq \left\{ \sum_{n=n_0+1}^{n_{2^k-k_2}} \chi_n = 0 \right\} \end{aligned} \quad (45)$$

where, (45) follows from the fact that  $k \geq k_1$  and hence  $\mathcal{O}' = r_{k_1} \bar{U} \subseteq r_k \bar{U}$ . It is also worth mentioning here that the proof of Lemma 5.8(a) holds irrespective of how the iterates are generated. Given that  $\omega \in E(k, n_0) \cap \left( \cap_{m=0}^{2^{k-k_2}-1} \{\tilde{\rho}_1^{(l)}(\omega, n_m, n_0, T_m) + \tilde{\rho}_2^{(l)}(\omega, n_m, n_0, T_m) < \epsilon_0\} \right)$ , along this sample path there has been a reset at  $n_0$  and at the next check performed at  $n_{2^k-k_2}$  there has been no reset. Hence the next check for reset is performed by Algorithm 1 at  $n_{2^{(k-k_2)}+1}$ . Again from arguments from Lemma 5.8(a), we get that,

$$\begin{aligned} E(k, n_0) \cap \left( \cap_{m=0}^{2^{(k-k_2)}+1-1} \{\tilde{\rho}_1^{(l)}(\omega, n_m, n_0, T_m) + \tilde{\rho}_2^{(l)}(\omega, n_m, n_0, T_m) < \epsilon_0\} \right) &\subseteq \{\omega \in \Omega : X_{2^{(k-k_2)}+1}(\omega) \in \mathcal{O}'\} \\ &\subseteq \left\{ \sum_{n=n_0+1}^{n_{2^{(k-k_2)}+1}} \chi_n = 0 \right\}. \end{aligned}$$

Repeating the above for the third reset check after  $n_0$  and so on, we obtain that,

$$E(k, n_0) \cap \left( \cap_{m \geq 0} \{\tilde{\rho}_1^{(l)}(\omega, n_m, n_0, T_m) + \tilde{\rho}_2^{(l)}(\omega, n_m, n_0, T_m) < \epsilon_0\} \right) \subseteq \left\{ \sum_{n=n_0+1}^{\infty} \chi_n = 0 \right\}. \quad (46)$$

**Step 7 (Bounding) :** Define  $\tilde{\mathcal{B}}_{-1}^{(l)} := E(k, n_0)$  and for every  $M \geq 1$  define,

$$\tilde{\mathcal{B}}_M^{(l)} := E(k, n_0) \cap \left( \cap_{m=0}^M \{\tilde{\rho}_1^{(l)}(\cdot, n_m, n_0, T_m) + \tilde{\rho}_2^{(l)}(\cdot, n_m, n_0, T_m) < \epsilon_0\} \right).$$

Note that as in Lemma 5.10, we can obtain  $l_0 \geq 1$ , such that for every  $m \geq 0$ , for every  $\omega \in \tilde{\mathcal{B}}_{m-1}^{(l_0)}$  we have that  $\tilde{\rho}_2^{(l_0)}(\omega, n_m, n_0, T_m) < \frac{\epsilon_0}{2}$ , since for every  $\omega \in \tilde{\mathcal{B}}_{m-1}^{(l_0)}$ ,  $X_{n_m}(\omega) \in \mathcal{O}'$  and whether or not a reset

check is performed at this index, we have that  $X_{n_m}(\omega) = X'_{n_m}(\omega)$ . Thus for such an  $l_0$ , mimicking the proof of Lemma 5.10, we obtain that,

$$\mathbb{P}\left(\cap_{m=0}^M \{\tilde{\rho}_1^{(l)}(\cdot, n_m, n_0, T_m) + \rho_2^{(l)}(\cdot, n_m, n_0, T_m) < \epsilon_0\} | \tilde{\mathcal{B}}_{-1}^{(l_0)}\right) \geq 1 - \sum_{m=0}^{\infty} \mathbb{P}\left(\tilde{\rho}_1^{(l_0)}(\cdot, n_m, n_0, T_m) \geq \frac{\epsilon_0}{2} | \tilde{\mathcal{B}}_{m-1}^{(l_0)}\right).$$

From Lemma 5.9, we have that for  $k > \max\{k_1, k_2, N'_0\}$ , for every  $n_0 \in J(k)$ ,

$$\mathbb{P}\left(\cap_{m=0}^M \{\tilde{\rho}_1^{(l)}(\cdot, n_m, n_0, T_m) + \rho_2^{(l)}(\cdot, n_m, n_0, T_m) < \epsilon_0\} | \tilde{\mathcal{B}}_{-1}^{(l_0)}\right) \geq 1 - \sum_{m=0}^{\infty} \mathbb{P}\left(\max_{n_m \leq j \leq n_{m+1}} \|\zeta_j - \zeta_{n_m}\| \geq \frac{\epsilon_0}{4K_0(T_u)} | \tilde{\mathcal{B}}_{m-1}^{(l_0)}\right). \quad (47)$$

**Step 8 (Noise bound)** Similar to item (a) in section 5.4, from Algorithm (1), we have that for every  $m \geq 0$ , for every  $n_m \leq n_{m+1} - 1$ ,  $\|X'_{j+1}\| \leq \|X'_j\|(1 + a(j)K) + a(j)K + a(j)\|M_{j+1}\|$  and since  $\|X'_{j+1}\| \leq \|X_{j+1}\|$ , we get that for every  $n_m \leq j \leq n_{m+1} - 1$ ,

$$\|X'_{j+1}\| \leq \|X'_j\|(1 + a(j)K) + a(j)K + a(j)\|M_{j+1}\|.$$

Now by arguments exactly same as those item (a) of section 5.4, we get that for every  $m \geq 0$ , for every  $n_m \leq j \leq n_{m+1} - 1$ ,  $\|X'_{j+1}\| \leq e^{2KT_u}(\|X_{n_m}\| + 2KT_u)$ . Now by using concentration inequality as in item (b) of section 5.4, we get that for every  $m \geq 0$ ,

$$\mathbb{P}\left(\max_{n_m \leq j \leq n_{m+1}} \|\zeta_j - \zeta_{n_m}\| \geq \frac{\epsilon_0}{4K_0(T_u)} | \tilde{\mathcal{B}}_{m-1}^{(l_0)}\right) \leq 2de^{-\tilde{K}/(b(n_m) - b(n_{m+1}))}. \quad (48)$$

Using (46), (47) and (48) we get that, for every  $k \geq \max\{k_1, k_2, N_0\}$  (where  $N_0$  is as defined in the proof of Theorem 3.2), for every  $n_0 \in J(k)$ ,

$$\mathbb{P}\left(\sum_{n=n_0+1}^{\infty} \chi_n = 0 \mid \sum_{n=1}^{n_0-1} \chi_n = k-1, \chi_{n_0} = 1\right) \geq 1 - 2d \sum_{m=0}^{\infty} e^{-\tilde{K}/(b(n_m) - b(n_{m+1}))} \geq 1 - 2de^{-\tilde{K}/b(n_0)}.$$

Substituting the above in (41) and using the fact that for  $n \leq n'$ ,  $b(n') \leq b(n)$ , we get that, for every  $k \geq \max\{k_1, k_2, N_0\}$ ,

$$\begin{aligned} \mathbb{P}\left(\sum_{n=1}^{\infty} \chi_n = k\right) &\geq (1 - 2de^{-\tilde{K}/b(k)}) \sum_{n_0=1}^{\infty} \mathbb{P}\left(\left\{\sum_{n=1}^{n_0-1} \chi_n = k-1\right\} \cap \{\chi_{n_0} = 1\}\right) \\ &= (1 - 2de^{-\tilde{K}/b(k)}) \mathbb{P}\left(\left\{\sum_{n=1}^{\infty} \chi_n \geq k\right\}\right) \end{aligned} \quad (49)$$

Substituting (49) in (37), we get that for every  $k \geq \max\{k_1, k_2, N_0\}$ ,

$$\mathbb{P}(G_k) \geq \mathbb{P}(G_{k-1}) + (1 - 2de^{-\tilde{K}/b(k)})\mathbb{P}(G_{k-1}^c) \geq 1 - 2de^{-\tilde{K}/b(k)}.$$

Letting  $k \rightarrow \infty$  in the above equation and using the fact that  $\mathbb{P}(G_{\infty}) = \lim_{k \rightarrow \infty} \mathbb{P}(G_k)$ , we get that  $\mathbb{P}(G_{\infty}) = 1$ .

## 7 Conclusions and directions for future work

We have extended the lock-in probability result (Theorem 3.2) in [6] to stochastic approximation schemes with set-valued drift functions which serves as an important tool for analyzing recursions when their stability is not guaranteed. The extension to set-valued map allows one to obtain lock-in probability for stochastic approximation schemes with measurable drift functions and schemes where the drift function itself possess a non-additive unknown noise component (see [3, Ch. 5.3]). Further using Theorem 3.2, in the presence of a locally attracting set for the mean field, we have provided an alternate condition for verification of convergence in the absence of stability guarantee which involves verifying whether the iterates are entering infinitely often, an open neighborhood of the attractor with a compact closure. In the presence of a globally attracting set our modified recursion as in Algorithm 1, converges almost surely to the globally attracting set, the proof of which relies on the method used to obtain the lock-in probability result.

In future we wish to consider other applications of the lock-in probability result such as sample complexity (see [3, Ch. 4.2]) and almost sure convergence under tightness of the iterates (see [7]). Another interesting direction, is to explore various additive noise models where the above result can be extended for the case of set-valued drift functions.

## A Definitions of some topological concepts

Let  $(\mathcal{M}, \Gamma)$  be a topological space.  $\{O_i\}_{i \in I}$  is an *covering* of  $\mathcal{M}$  if, for every  $i \in I$ ,  $O_i \subseteq \mathcal{M}$  and  $\cup_{i \in I} O_i = \mathcal{M}$ . Further a covering  $\{O_i\}_{i \in I}$  is said to be *locally finite* if for every  $p \in \mathcal{M}$ , there exists an  $O \in \Gamma$ , such  $O_i \cap O \neq \emptyset$  for only finitely many  $i \in I$ . Given any two coverings  $\mathcal{C} := \{O_i\}_{i \in I}$  and  $\mathcal{C}' := \{O'_j\}_{j \in J}$ ,  $\mathcal{C}'$  is said to be a *refinement* of  $\mathcal{C}$  if, for every  $i \in I$ , there exists a  $j \in J$  such that,  $O_i \subseteq O'_j$ .  $\mathcal{C}'$  is said to be a *locally finite refinement* of  $\mathcal{C}$ , if  $\mathcal{C}'$  is a refinement of  $\mathcal{C}$  and is locally finite. The topological space  $(\mathcal{M}, \Gamma)$  is *paracompact* if it is a Hausdorff space and if every open covering has a locally finite open refinement.

A family of functions  $\{\psi_i\}_{i \in I}$  is called a *locally Lipschitz partition of unity* if for all  $i \in I$ ,

- $\psi_i$  is locally Lipschitz continuous and non negative,
- the supports of  $\psi_i$ , defined as  $\overline{\{p \in \mathcal{M} : \psi_i(p) \neq 0\}}$ , are a closed locally finite covering of  $\mathcal{M}$ ,
- for each  $p \in \mathcal{M}$ ,  $\sum_{i \in I} \psi_i(p) = 1$ .

A partition of unity  $\{\psi_i\}_{i \in I}$  is said to be *subordinated* to the covering  $\{O_i\}_{i \in I}$ , if for every  $i \in I$ ,  $\{p \in \mathcal{M} : \psi_i(p) \neq 0\} \subseteq O_i$ .

## B Proof of Lemma 5.9

Fix  $l \geq 1$  and  $T_u \geq T_A$ . Fix  $n_0 \geq 1$  and  $\{n_m\}_{m \geq 0}$  as in **Definition 5**. Fix  $m \geq 0$ . Let  $\omega \in \{\omega \in \Omega : \rho_1^{(l)}(\omega, n_m, T_m) \geq \frac{\epsilon_0}{2}\} \cap \mathcal{B}_{m-1}^{(l)}$ . Then, by Lemma 5.8(a) we have that,  $X_{n_m}(\omega) \in \mathcal{O}'$ . By (9) and **Definition 3**, we have that for every  $0 \leq t \leq T_m$ , there exists  $n_m \leq k \leq n_{m+1} - 1$  such that  $t \in [t(k), t(k+1)]$ , and

$$\bar{X}(\omega, t + t(n_m)) = \alpha X_{n_k}(\omega) + (1 - \alpha) X_{n_{k+1}}(\omega), \quad (50)$$

for some  $\alpha \in [0, 1]$  and

$$\tilde{x}^{(l)}(t; n_m, T_m, \omega) = X_{n_m}(\omega) + \int_0^q f^{(l)}(\tilde{x}^{(l)}(q; n_m, T_m, \omega), u(q; n_m, T_m, \omega)) dq. \quad (51)$$

Therefore for any  $t \in [0, T_m]$ ,

$$\begin{aligned} \|\bar{X}(\omega, t + t(n_m)) - \tilde{x}^{(l)}(t; n_m, T_m, \omega)\| &\leq \alpha \|X_k(\omega) - \tilde{X}^{(l)}(t(k); n_m, T_m, \omega)\| \\ &\quad + \alpha \|\tilde{X}^{(l)}(t(k); n_m, T_m, \omega) - \tilde{X}^{(l)}(t; n_m, T_m, \omega)\| \\ &\quad + (1 - \alpha) \|\tilde{X}^{(l)}(t(k+1); n_m, T_m, \omega) - \tilde{X}^{(l)}(t; n_m, T_m, \omega)\| \\ &\quad + (1 - \alpha) \|X_{k+1}(\omega) - \tilde{X}^{(l)}(t(k+1); n_m, T_m, \omega)\| \end{aligned} \quad (52)$$

Now the aim is to provide an upper bound for each terms in the *R.H.S.* of the above inequality which is independent of  $m$  and  $t$ . In order to apply [3, Ch. 2, Lemma 1] the only additional condition needed is the Lipschitz continuity of  $f^{(l)}(\cdot, \cdot)$  uniformly over  $u$  and this is obtained using local Lipschitz continuity as follows. Let  $M_1 > 0$  be such that, for every  $x \in \mathcal{O}'$ ,  $\|x\| \leq M_1$ . By Lemma 5.7(b) and item (a) in section 5.4, we have that for  $r := \max\{C_1(\bar{\mathcal{O}}', T_u, l), e^{2KT_u}(M_1 + 2KT_u)\}$ ,

$$\sup_{t \in [0, T_m]} \|\bar{X}(\omega, t + t(n_m))\| \leq r \text{ and } \sup_{t \in [0, T_u]} \|\tilde{x}^{(l)}(t; n_m, T_m, \omega)\| \leq r.$$

Further by Lemma 5.7(c), we know that there exists  $L(r, l) > 0$ , such that for every  $x, x' \in rU$ , for every  $t \in [0, T_m]$

$$\|f^{(l)}(x, u(t; n_m, T_m, \omega)) - f^{(l)}(x', u(t; n_m, T_m, \omega))\| \leq L(r, l) \|x - x'\|.$$

The rest of the bounding procedure is exactly the same as [3, Ch. 2, Lemma 1]. We obtain that, for every  $t \in [0, T_m]$ ,

$$\begin{aligned} \sup_{t \in [0, T_m]} \|\bar{X}(\omega, t + t(n_m)) - \tilde{x}^{(l)}(t; n_m, T_m, \omega)\| &\leq (M_1 + KT_u) e^{2L(r, l)T_u} L(r, l) \sum_{j \geq 0} a(n_0 + j)^2 \\ &\quad + e^{L(r, l)T_u} \max_{n_m \leq j \leq n_{m+1}} \|\zeta_j(\omega) - \zeta_{n_m}(\omega)\| \\ &\quad + (M_1 + KT_u) e^{L(r, l)T_u} a(n_0). \end{aligned}$$



Now set  $N'_0$  such that  $(M_1 + KT_u)e^{2L(r,l)T_u}L(r,l)\sum_{j \geq 0} a(n_0+j)^2 + (M_1 + KT_u)e^{L(r,l)T_u}a(n_0) < \frac{\epsilon_0}{2}$  and define  $K_0(T_u) := e^{L(r,l)T_u}$ . Then, for every  $n_0 \geq N'_0$ , we have that  $\omega \in \{\omega \in \Omega : \max_{n_m \leq j \leq n_{m+1}} \|\zeta_j(\omega) - \zeta_{n_m}(\omega)\| \geq \frac{\epsilon_0}{4K_0(T_u)}\} \cap \mathcal{B}_{m-1}^{(l)}$ , from which Lemma 5.9 follows.

## References

- [1] M. Benaïm, “Dynamics of stochastic approximation algorithms,” in *Seminaire de probabilités XXXIII*. Springer, 1999, pp. 1–68.
- [2] M. Benaïm, J. Hofbauer, and S. Sorin, “Stochastic approximations and differential inclusions,” *SIAM Journal on Control and Optimization*, vol. 44, no. 1, pp. 328–348, 2005.
- [3] V. S. Borkar, *Stochastic approximation : A dynamical systems viewpoint*. Cambridge University Press, 2008.
- [4] V. S. Borkar and S. P. Meyn, “The ODE method for convergence of stochastic approximation and reinforcement learning,” *SIAM Journal on Control and Optimization*, vol. 38, no. 2, pp. 447–469, 2000.
- [5] A. Ramaswamy and S. Bhatnagar, “A generalization of the Borkar-Meyn theorem for stochastic recursive inclusions,” *arXiv preprint arXiv:1502.01953*, 2015.
- [6] V. S. Borkar, “On the lock-in probability of stochastic approximation,” *Combinatorics, Probability and Computing*, vol. 11, no. 1, p. 11–20, Jan 2002.
- [7] S. Kamal, “On the convergence, lock-in probability, and sample complexity of stochastic approximation,” *SIAM Journal on Control and Optimization*, vol. 48, no. 8, pp. 5178–5192, 2010.
- [8] H.-F. Chen, L. Guo, and A.-J. Gao, “Convergence and robustness of the Robbins-Monro algorithm truncated at randomly varying bounds,” *Stochastic Processes and their Applications*, vol. 27, pp. 217–231, 1987.
- [9] C. HANFU and Z. YUNMIN, “Stochastic approximation procedures with randomly varying truncations,” *Science in China Series A-Mathematics, Physics, Astronomy & Technological Science*, vol. 29, no. 9, pp. 914–926, 1986.
- [10] G. Fort, E. Moulines, A. Schreck, and M. Vihola, “Convergence of Markovian stochastic approximation with discontinuous dynamics,” *SIAM Journal on Control and Optimization*, vol. 54, no. 2, pp. 866–893, 2016.
- [11] C. Andrieu, M. Vihola *et al.*, “Markovian stochastic approximation with expanding projections,” *Bernoulli*, vol. 20, no. 2, pp. 545–585, 2014.
- [12] J.-P. Aubin and A. Cellina, *Differential inclusions: Set-valued maps and viability theory*. Springer Science & Business Media, 2012, vol. 264.
- [13] S. Li, Y. Ogura, and V. Kreinovich, *Limit theorems and applications of set-valued and fuzzy set-valued random variables*. Springer Science & Business Media, 2013, vol. 43.
- [14] V. Yaji and S. Bhatnagar, “Stochastic recursive inclusions with non-additive iterate-dependent Markov noise,” *arXiv preprint arXiv:1607.04735*, 2016.
- [15] J.-P. Aubin and H. Frankowska, *Set-valued analysis*. Springer Science & Business Media, 2009.
- [16] D. O’Regan, *Existence theory for nonlinear ordinary differential equations*. Springer Science & Business Media, 1997, vol. 398.